December 2017

The Evolutionary Diversity of Uracil DNA Glycosylase Superfamily

Jing Li
Clemson University, lijing521345@outlook.com

THE EVOLUTIONARY DIVERSITY OF URACIL DNA GLYCOSYLASE SUPERFAMILY

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biochemistry and Molecular Biology

---

by
Jing Li
December 2017

Accepted by:
Dr. Weiguo Cao, Committee Chair
Dr. Alex Feltus
Dr. Cheryl Ingram-Smith
Dr. Jeremy Tzeng

ABSTRACT

Uracil DNA glycosylase (UDG) is a crucial member in the base excision (BER) pathway that is able to specially recognize and cleave the deaminated DNA bases, including uracil (U), hypoxanthine (inosine, I), xanthine (X) and oxanine (O). Currently, based on the sequence similarity of 3 functional motifs, the UDG superfamily is divided into 6 families. Each family has evolved distinct substrate specificity and properties. In this thesis, I broadened the UDG superfamily by characterization of three new groups of enzymes. In chapter 2, we identified a new subgroup of enzyme in family 3 SMUG1 from *Listeria Innocua*. This newly found SMUG1-like enzyme has distinct catalytic residues and exhibits strong preference on single-stranded DNA substrates. In chapter 3, we extensively investigated an untraditional family 1 UNG enzyme from *Nitratifractor salsuginis* (Nsa UNG). This enzyme is able to form a unique salt bridge network with uracil-containing DNA. In addition, this untraditional family 1 UNG can't be inhibited by uracil DNA glycosylase inhibitor (Ugi). In chapter 4, a potential evolutionary immediate between family 1 UNG and family 4 UDGa was isolated from *Janthinobacterium agaricidamnosum* (Jag UNG). In the functional motifs, Jag UNG has evolved family 1 UNG unique features, but still keeps some features of family 4 UDGa. Through site-directed mutagenesis, molecular modeling and biophysical analysis, we estimated that QD in family 1 UNG might be coevolved from EG in family 4 UDGa. In addition, we found another two important sites (A82E and L245H) that may have special meaning in the evolutionary history. All these work reveal the evolutionary diversity of UDG superfamily.

DEDICATION

This dissertation is dedicated to my parents, husband and family.

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Weiguo Cao who is supporting, helping and advising me during my Ph.D. years in Clemson University.

I also sincerely thank my committee members (Dr. Alex Feltus, Dr. Cheryl Ingram-Smith, Dr. Jeremy Tzeng) for their support, assistance and advice to pursue my Ph.D. degree.

I also would like to thank all previous lab members (Dr. Guang-chen Fang, Dr. Bo Xia, Dr. Dong-Hoon Lee) and all current lab members (Dr. Ye Yang, Chenyan Chang, Anqi Wei) and all undergraduate students involved in my research projects (Alyssa Shearer) for their help and discussions.

I would like to express thanks to our collaborators, Dr. Wei Xie from Sun Yat-sen University, for his help on crystal structure determination. I also give thanks Dr. Liangjiang Wang and Jose Guevara helped us performed the mutual information analysis of UDG superfamily. I also would like to thank to Dr. Min Cao from biological department for providing the *Listeria Inoccua* strain. I would like to thank Dr. Emil Alex from physics department for training of biophysical skills.

I want to say thanks to Dr. Tzeng's family for providing a home when I arrived Clemson several years ago. I also give thanks Dr. Cao's family for accompany of each spring festival. I am appreciated the support and accompany of my parents and husband. I especially thank my little daughter for coming to my life.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

Figure Page

ABBREVIATIONS

AAG, Alkyladenine DNA glycosylase

AER, Alternative excision repair

ADA, Adenosine deaminase

AID, Activation-induced deaminase

AlkA, 3-methyladenine DNA glycosylase II

APE1, Human AP endonuclease

AP site, Apurinic/apyrimidinic site

BER, Base excision repair

CPDs, cyclobutane pyrimidine dimers

CSR, class switch recombination

DSB, DNA double-strand break

DSBR, DNA double-strand break repair

Endo IV, Endonuclease IV

Exo III, Exonuclease III

$FADH^-$, Flavin adenine dinucleotide

FMN, Flavin mononucleotide

GG-NER, Global genomic nucleotide excision repair

GDA, Guanine deaminase

HDG, Hypoxanthine DNA glycosylase

HR, Homologous recombination

MAG, 3-methyledenine-DNA-glycosylase

Abbreviations (Continued)

MMR, Methyl-directed mismatch repair

MUG, Mismatch-specific uracil DNA glycosylase

NER, Nucleotide excision repair

NHEJ, Non-homologous end joining

NRAMP1, Natural resistance-associated macrophage protein 1

NNAL, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol

NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone

NNN, N'-nitrosonornicotine

NTH, Endonuclease iii

NTHL, Endonuclease iii-like

NEIL, Endonuclease IIIV-like

OGG1, 8-Oxoguanine DNA glycosylase

PARP1, Poly (ADP-ribose) polymerase 1

PARG, Poly (ADP-ribose) glycohydrolase

PNKP, polynucleotide kinase 3'-phosphatase

RAG, Recombinase-activating gene

RNS, Reactive nitrogen species

ROS, Reactive oxygen species

RPA, Replication protein A

SHM, Somatic hypermutation

SMUG1, Single-stranded selective monofunctional uracil-DNA glycosylase

Abbreviations (Continued)

SSA, Single strand annealing

SSB, Single strand break

SSBP, Single strand binding protein

SSBR, Single strand break repair

TC-NER, transcription-coupled nucleotide excision repair

TDG, Thymine DNA glycosylase

TDP1, tyrosyl-DNA phosphodiesterase 1

TRCF, Transcription-repair coupling factor

TOP2, Topoisomerase 2

UNG, Uracil N-glycosylase

UDG, Uracil DNA glycosylase

XDG, Xanthine DNA glycosylase

CHAPTER ONE

DNA DAMAGE, REPAIR AND URACIL DNA GLYCOSYLASE

I. Introduction

As the genetic information carrier, DNA is essential for each living cell. However, it is vulnerable to the attack from endogenous and exogenous reactive oxygen species (ROS), reactive nitrogen species (RNS) and alkylation agents. Especially, DNA bases are continually suffering oxidation, deamination, alkylation and depurination. All of these base damages are highly mutagenic if not repaired. In addition, single strand break (SSB) and double strand break (DSB) are more dangerous for the integrity of genetic information. To protect DNA from all kinds of damage, organisms have evolved diverse DNA repair systems. In the following paragraphs, different type of DNA base damage, SSB and DSB will be discussed. The respective DNA repair strategies including base excision repair (BER), nucleotide excision repair (NER), mis-matched repair (MMR), photoreactivation, single strand break repair (SSBR) and double strand break repair (DSBR) will be elaborated. Especially, the crucial enzymes in each pathway will be described in detail.

II. DNA damage

DNA bases and DNA strands are suffering different kinds of damages due to the attacks from environmental and endogenous reagents. Here, I discussed DNA base lesions, single strand break (SSB) and double strand break as below (DSB).

A. DNA base legions

1. DNA base oxidation

DNA base oxidation is mainly caused by the reactive oxygen species (ROS) that are spontaneously produced in cellular metabolisms or induced by exogenous stress such as UV radiation, ironizing radiation and heat exposure [1]. To date, more than 20 kinds of oxidative DNA base lesions have been identified (Fig. 1.1) [2], however, only a few of them are well studied. For instance, 8-hydroxyguanine (8-OH-Gua) is extensively investigated as larger percentage occurs *in vivo* and it is easy to detect [3, 4]. Once 8-OH-Gua lesion is introduced into DNA, it might lead to mis-reading by DNA polymerase [5]. Moreover, 8-OH-Gua/C base pair will cause a G/C to T/A mutation and misincorporation of 8-OH-Gua opposite of T will produce an A to C mutation [6]. Numerous studies showed that the elevated level of 8-OH-Gua is associated with many types of cancers and diseases including acute lymphoblastic leukemia, breast cancer, cervical cancer, colorectal cancer, gastric cancer, renal cell carcinoma, lung cancer and so on [7]. The increased level of other oxidative lesions including, 5-hydroxycytosine, 5-hydroxyuracil, 5-hydroxymethyluracil, 5-hydroxy-5-methyl-hydantoin, 5-hydroxy-hydantoin, thymine glycol, 8-hydroxyadenine, and 2-hydroxyadenine are also related to human diseases [8-10].

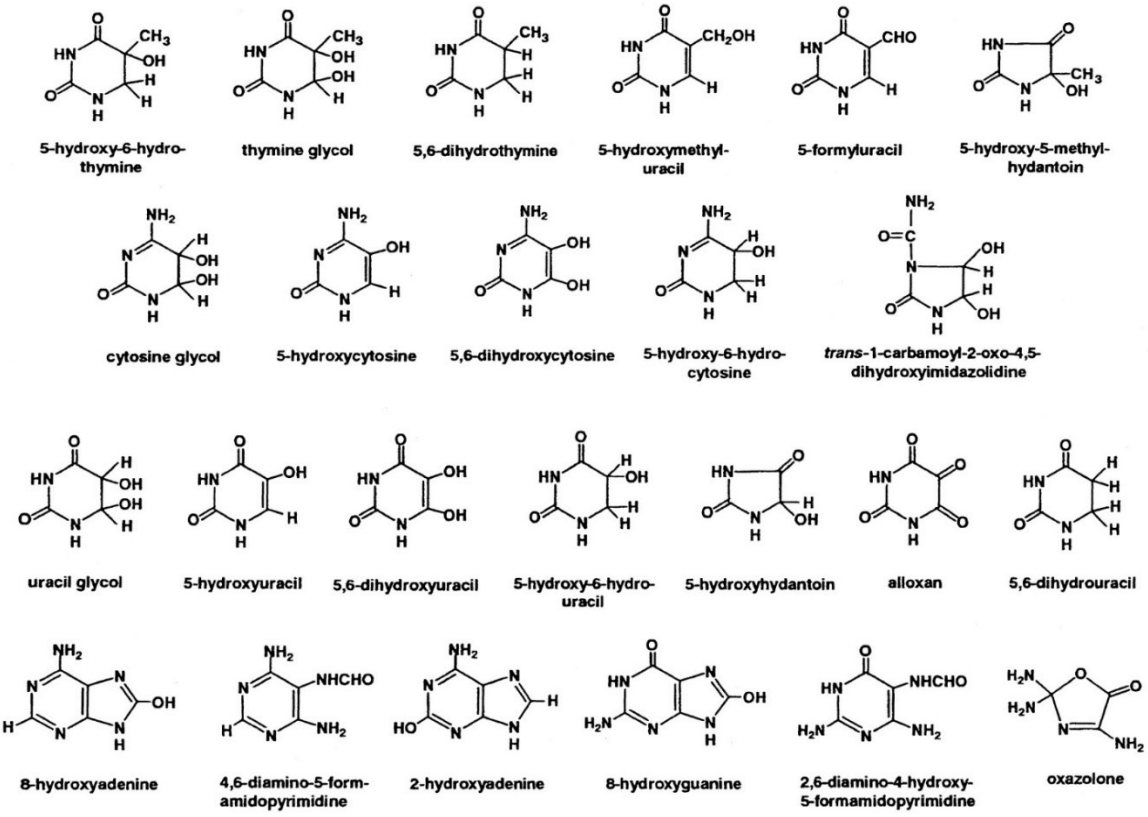


Figure 1. 1 Chemical structures of major oxidative DNA base lesions. The picture was adopted from [7] with permission.

2. DNA Alkylation

DNA alkylation lesions are inevitable as the result of the ubiquitous alkylating agents. S-Adenosylmethionine, a natural methyl-donor *in vivo*, is an endogenous alkylating agent. It is able to transfer its methyl group to guanine and produce 7-meG, 3-meG and O^6-meG [11, 12]. Meanwhile, large amount of alkylating agents are present in environment, including chloromethane, bromomethane and lodomethane [13]. Besides, the nitroso compounds in tobacco smoke, such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and N'-nitrosonornicotine (NNN) are potential DNA alkylating agents. They are able to increase

the level of 7-meG, O8-meG, O4-meT and the larger pyridyloxobutyl G adducts (Fig. 1.2) [14]. All of the alkylation lesions in DNA are mutagenic, genotoxic and cytotoxic. Given such deadly effects, several alkylation agents (such as temozolomide and carmustin, lomusine and fotemustine) have been used in therapy as anti-cancers drugs.

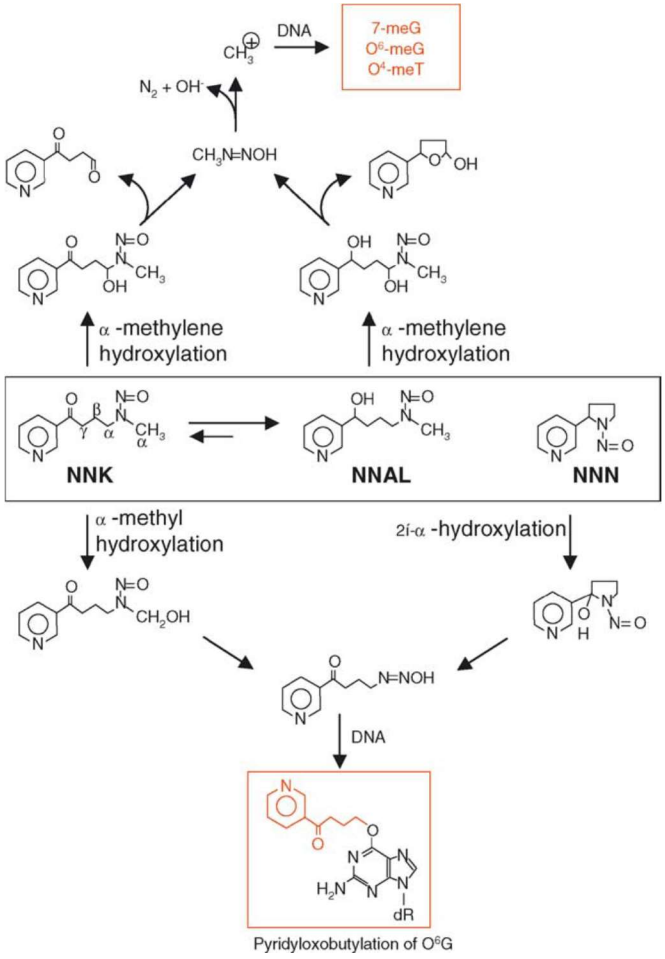


Figure 1. 2 Overview of the formation of DNA adducts from tobacco-specific NNK, NNAL and NNN. NNK, NNAL and NNN are going through α-methylene hydroxylation or 2 α-hydroxylation and finally give rise to 7-meG, O8-meG, O4-meT and the larger pyridyloxobutyl G adducts. The picture was reused from [14] with permission.

3. DNA Depurination

DNA depurination refers to the hydrolysis of N-glycosidic bond of dG and dA in DNA. The resulting apyrimidinic (AP) sites are very harmful for the integrity of genomic DNA. The frequency of DNA depurination is very high in mammalian cells, around 2000-10000 purine bases are lost in each human cell every day [15]. Besides the spontaneous depurination, AP sites are naturally produced in the BER pathway as the cleavage of damaged or mismatched bases by DNA glycosylase [16]. AP sites are very cytotoxic and mutagenic. Firstly, by stalling DNA and RNA polymerase during DNA replication and transcription, AP site may lead to cell death [17]. Secondly, even though DNA polymerase is not blocked by AP site during DNA replication, the AP-site containing DNA strand can't be used as a correct template for replication. A random base will be inserted opposite of an AP site, therefore a base substitution might be induced in the offspring DNA strands [18]. Thirdly, cleavage of AP site by AP endonuclease is a possible source of single-stranded break (SSB) that is not suitable for DNA replication and even possibly form more severe double-stranded break (DSB) [19].

4. DNA base damage by UV radiation

Ultraviolet (UV) is a natural part of solar energy. Therefore, DNA bases are inevitably attacked by the UV radiation. Especially, UV-b (280 nm-315 nm) is the main cause of thymine dimers, thymine-cytidine dimers and cytidine-cytidine dimers. All those pyrimidine dimers including cyclobutane pyrimidine dimers (CPDs), pyrimidine 6-4 pyrimidone photoproducts, and their Dewar isomers are highly mutagenic and cytotoxic [20]. Base dimers largely alter the DNA conformation and significantly affect the

stability and functions of DNA [21-23]. Base dimers can be directly spited by photolyase in photoreactivation pathway or excised in the nucleotide excision repair (NER) pathway that will be discussed later.

B. Single strand break (SSB)

The discontinuity of one of the double DNA strands is termed as single strand break (SSB). As one of the most frequent DNA damages, up to 10, 000 SSBs were detected in each cell every day [24]. Both endogenous and exogenous ROS trigger SSBs by attacking the sugar phosphate backbone of DNA. In the BER pathway, the cleavage of AP sites by AP endonuclease may give rise to SSBs if the following DNA polymerase doesn't work efficiently [25, 26]. DNA topoisomerase 1 create a nick in one of the DNA strands to relax supercoils that is also a potential source of SSB [27]. Unrepaired SSBs could threaten the life of cells by blocking the DNA replication and transcription [28-31]. Evidences showed that two genetic diseases, ataxia-oculomotor apraxia 1 and spinocerebellar ataxia with axonal neuropathy 1, are associated with SSBs repair deficiency [32].

C. Double strand break (DSB)

Double strand break (DSB) is one common type of DNA damage under spontaneous or radiation-induced oxidative stress [33]. There are an estimated 10 DSBs in each cell every day [34]. The intermediates of V (D) J recombination, class-switch recombination, and meiotic recombination are naturally potential sources of DSBs [35]. Dysfunction of some enzymes, such as DNA topoisomerases 2(TOP2), activation-induced deaminase (AID) and the recombinase-activating gene (RAG) endonuclease are

contributing to the increase of DSBs [36, 37]. Also, as mentioned before, unrepaired SSBs are offering DSBs after DNA replication [28]. The consequences of DSBs, including mutations, deletions, translocations and reversions are very deadly to the cells [38]. Disruption or deficiency of DSB repair is related to a variety of human genetic diseases and cancers, including nijmegen breakage syndrome, seckel syndrome, ataxia telangiectagia like disorder, radiosensitive severe combined immunodeficiency, breast and ovarian cancer [39].

III. DNA repair

In order to counter various DNA damage and protect the genetic information carrier, organisms have evolved diverse DNA repair mechanisms. In the following paragraphs, different DNA repair pathways, including base excision repair (BER), photoreactivation, nucleotide excision repair (NER), mis-matched repair (MMR), single strand break repair (SSBR) and double strand break repair (DSBR) will be introduced.

A. Base Excision Repair (BER)

Base excision repair (BER) pathway is highly conserved in both prokaryotes and eukaryotes. Through the BER pathway, oxidative, alkylated and deaminated DNA lesions will be removed. In the BER pathway, DNA glycosylase is the initial enzyme that detects the damaged bases and removes them by cleaving the N-glycosidic bond. The resulting AP site will be eliminated by AP endonuclease. Then, DNA polymerase and ligase will seal the nick left in the DNA strand [40]. As a crucial enzyme in BER pathway, DNA glycosylase draws lots of attention from scientists. So far, different kinds of DNA glycosylase have been identified, each of which has distinct functions.

1. 8-Oxoguanine glycosylase (OGG1)

8-Oxoguanine glycosylase was firstly identified from yeast and mammal with removal activity of 8-oxoG from 8-oxoG. C base pair [41, 42]. Later, a new activity toward another oxidative base 2, 6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) was discovered [43, 44]. Deletion of OGG1 from mice give rise to 8-oxoG level and susceptibility to obesity and metabolic dysfunction [45, 46]. Human OGG1 mutations are proved to be associated with non-small cell lung cancer and childhood acute lymphoblastic leukemia [47, 48].

2. MutY DNA glycosylase

MutY is another well characterized bacterial glycosylase contributing to defense 8-oxoG. Instead of cleaving 8-oxoG itself, MutY identifies and excises adenine from A.8-oxoG, a mis-pair due to the miscorportion of 8-oxoG [49]. Besides, MutY is able to remove A from other mismatched base pairs such as A.G [50]. The homolog of MutY in mammal, MUTYH, cleaves A opposite 8-oxoG, G, or C [49]. MUTYH gene mutations are associated with a hereditary colorectal cancer syndrome termed MUTYH associated polyposis [51].

3. 3-Methyladenine DNA glycosylase (AlkA)

Alkyladenine-DNA glycosylase (AlkA) are representative of *E. coli* 3-methyladenine-DNA glycosylase (AlkA), yeast 3-methyladenine-DNA glycosylase (MAG) and human 3-methyladenine DNA glycosylase (AAG). They are bifunctional to variety of alkylated lesions, including 3-methyladenine, 7-methylguanine, 3-methylguanine, O^2-methylthymine, O^2-methylcytosine, some oxidative lesions, 8-

oxoguanine, 5-hydroxymethyl uracil, 5-formyluracil and the deaminated lesion hypoxanthine [52-55]. In *E.coli* cells, AlkA is up-regulated by exogenous alkylated agents. In absence of AlkA, 7-methylguauosine level, mutation rate and SOS response are elevated in *E.coli* [56]. MAG deficient *Saccharomyces cerevisiae* cells increased susceptibility to methylation-induced recombination [57].

4. AP endonuclease

AP endonuclease is an indispensable enzyme in BER pathway that not only cleaves AP sites and some oxidative DNA lesions but also shows 3'phosphodiesterase activity and 3' to 5' exonuclease activity [58]. Currently, two families of AP endonuclease in bacteria are identified: endonuclease IV (Endo IV) and exonuclease III (Exo III). Endo IV are present only in some evolutionarily lower organisms including *E.coli*, *S. cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Danio rerio and Xenopus tropicalis* [59-61]. However, Exo III exists in both lower organisms and higher organisms, including plants and mammals. Catalytically, Endo IV enzymes relied on three Zn^{2+} ions and a hydroxide to cleave the phosphodiester bond, whereas Exo III members require Mg^{2+} [62]. Physiologically, Endo IV from both *S. cerevisiae* and *C. elegans* play an important role in DNA repair and maintaining the genome stability [60, 63]. The Exo III homolog in human, APE1 has complex functions *in vivo*. It is not only working as a DNA repair enzyme in BER pathway but also acting as a coactivator of various transcription factors in redox signaling. Given such important and complicated functions, APE1 mutations are associated with many diseases and cancers [64].

5. Nth (Endonuclease iii) and Nei (Endonuclease VIII)

Nth (Endonuclease iii) and Nei (Endonuclease VIII) are two DNA glycosylase with AP-lyase activity in bacteria and mammalians. Both of them are able to identify and remove diverse of oxidative DNA lesions, including thymine glycol, 8-oxoG, formamidopyrimidine, 5-hydroxycytosine, uracil glycol and 5-hydroxyuracil. Even though they have similar activities *in vitro*, a genetic assay showed that the spontaneous mutations are mainly repaired by Nth [65]. In *Mycobacterium smegmatis*, both Nth and Nei are responsible for the decrease of spontaneous mutation frequency [66]. In human, the homologs of Nth and Nei, NTHL1, NEIL1 and NEIL2 also work as essential DNA repair enzymes to defense the oxidative damages [67, 68]. Their variations are considered as a possible cause of the susceptibility of colorectal cancer [69, 70].

B. Photoreactivation

UV-induced DNA base dimers could be directly split by visible light dependent photoreactivation pathway. In this pathway, the photolyase consists two essential cofactors: light harvesting cofactor and catalytic cofactor. Light harvesting cofactor is able to absorb the visible light and transfer the energy to catalytic cofactor that split the dimers by transferring the light energy to them [71]. To date, few light harvesting cofactors have been identified, including 5, 10-methenyltetrahydrofolate, 8-hydroxy-5-deaza-riboflavin, and flavin mononucleotide (FMN) [72-74]. Deprotonated reduced flavin adenine dinucleotide ($FADH^-$) is the universal catalytic cofactor in all photolyase [75].

C. Nucleotide excision repair (NER)

Nucleotide excision repair (NER) pathway mainly participates in the repair of UV-induced DNA damages and this pathway is universal in prokaryotes and eukaryotes. In prokaryotes, depending whether RNA polymerase stalled in the DNA damaged lesion, NER pathway is divided into global genomic NER (GG-NER) and transcription coupled NER (TC-NER). In the GG-NER pathway, RNA polymerase is not stalled in the DNA lesion. First of all, the complex of UvrA and UvrB detects the damaged lesion and confirms the lesion by separating the two DNA strands. Then, UvrA is released and the nuclease UvrC is recruited by UvrB. Next, UvrB cleaves the phosphodiester bonds 8 nucleotides upstream of the damaged site and UvrC excises 4-5 nucleotides downstream of the damaged site. The helicase UvrD releases the 12-13 DNA lesion-containing oligonucleotides. Finally, the resulting DNA gap is filled and sealed by DNA polymerase and ligase.

If RNA polymerase is stalled in the DNA lesion, the transcriptional-repair coupling factor (TRCF) needs to remove the RNA polymerase and recruit UvrAB complex to initiate the following NER pathway, that is termed as transcription coupled NER (TC-NER) [76]. The main steps of eukaryotic NER are similar but more complicated than prokaryotes. More factors are necessary to process the NER pathway in eukaryotes. XPC/HR23B complex or DDBs are responsible for the DNA damage detection; CSB, CSA and XAB2 specifically participate in TC-NER pathway; the helicase TFIIH separates the DNA duplex. RPA is single stranded binding protein that is charged to open DNA duplex formation and stabilize the unwinded single stranded DNA; XPG and XPF/ERCC1 are endonucleases for the cleavage of phosphodiester bonds;

POLδ, POLε, POLκ and LIG1 are used to fill the gap [77]. Alternations in the key component of NER in human may cause severe UV-sensitive diseases such as xeroderma pigmentosum, cockayne syndrome, trichothiodystrophy and combined NER disorders. [77].

D. Mis-matched repair (MMR)

Mis-matched repair (MMR) is used to correct the mismatched and deletion/insertion base pairs introduced by DNA replication. Therefore, MMR is usually coupled with DNA replication. MMR is highly conserved in prokaryotes and eukaryotes. In *E.coli*, MutS identifies the mis-pairs and recruits MutL. The complex of MutS/MutL activates MutH endonuclease to create a nick in the unmethylated strand as in Fig. 1.3. Then the helicase Uvr D unwinds the double strand DNA and an exonuclease degrades the newly synthesized strand, leaving a gapped DNA strand. Later, DNA polymerase and ligase complete the rest of MMR. In eukaryotes, different enzymes are involved in MMR pathway. Firstly, MSH2-MSH6 recognize the mismatches. Secondly, the newly synthesized error-containing DNA strand is degraded by MSH2-MSH6, MLH1-PMS2 and EXOI. Thirdly, DNA polymerase resynthesizes the offspring DNA strand and DNA ligase completes the final step of MMR. In both prokaryotic and eukaryotic MMR, the single-strand binding protein (SSBP) and single-stranded DNA binding heterotrimer are essential in protecting the other single strand as the DNA replication template [78, 79]. Researchers found that mutations of the key enzymes (such as MLH1 and MSH2) in human MMR pathway, are associated with hereditary nonpolyposis colorectal cancer [80].

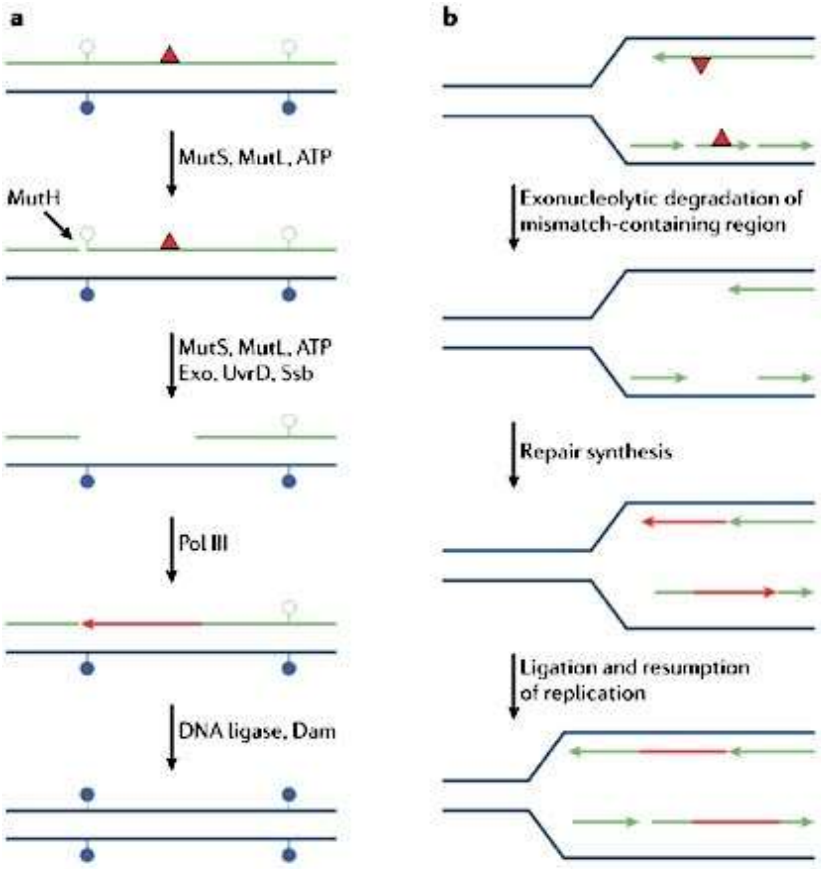


Figure 1. 3 DNA Mismatch Repair in Prokaryote and Eukaryote Systems. A. MMR in prokaryotes; B, MMR in Eukaryotes. The picture was reproduced from [79]with permission.

E. Single strand break repair (SSBR)

Based on the endogenous sources of single strand break (SSB), there are three sub-pathways of single strand break repair (SSBR) (Fig. 1.4). (1) In the BER pathway, when a SSB is created by AP endonuclease, a number of enzymes, including APE1, DNA polymerase (Pol) β, polynucleotide kinase 3'-phosphatase (PNKP) and aprataxin, are involved in DNA end processing. They work together to restore the 3'-hydroxyl (3'-OH) and 5'-phosphate groups for the following gap filling and DNA ligation that will be done by DNA polymerase and ligase. (2) For the sugar damaged (direct) SSBs, an

activated poly (ADP-ribose) polymerase 1 (PARP1) firstly identified the damage and then recovered to the pre-activated state by Poly (ADP-ribose) glycohydrolase (PARG). Later, APE1, PNKP and APTX participate the end processing followed by complicated gap filling. Finally, DNA ligase will finish the DNA repair. (3) Top1-linked SSB doesn't need PARP1 to be detected, tyrosyl-DNA phosphodiesterase 1 (TDP1) and PNKP directly process DNA ends. Next, they undergo a similar gap filling and ligation process to complete the DNA repair [32].

SSBR are very complicated but well organized and regulated. XRCC1 is reported to have an important role in stabilizing and stimulating other factors [81]. Evidences indicated that defect of some key components in SSBR is linked to some human diseases. For instance, alterations in APTX result in ataxia-oculomotor apraxia 1 and TDP1 mutations may cause spinocerebellar ataxia with axonal neuropathy 1. Both of those two diseases are genetic and characterized with chromosomal instability and cancer predisposition [82, 83].

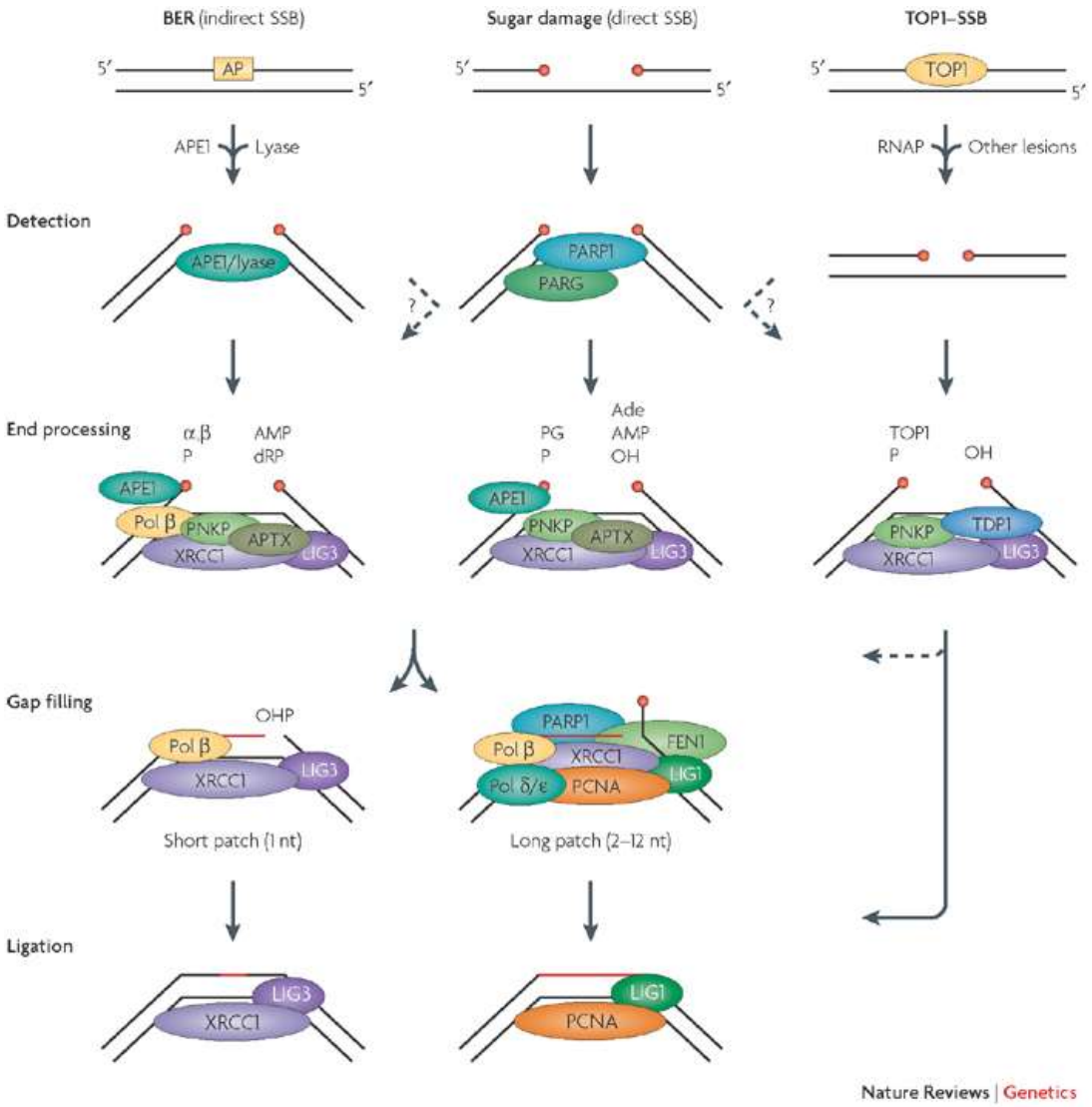


Figure 1. 4 Three sub-pathways of single strand break repair. A. indirect SSB; B. direct SSB; C TOP1-SSB. This picture was reused from [32] with permission.

F. Double strand break repair (DSBR)

Three mechanisms have evolved to protect DNA from the extremely genotoxic double strand breaks (DSBs): homologous recombination (HR) and nonhomologous DNA end joining (NHEJ) and single-strand annealing (SSA) (Fig. 1.5) [84]. All three DSBRs are conserved in prokaryotes and eukaryotes. However, they contribute differently in different organisms. It is established that HR has a major repair role in

prokaryotes and lower eukaryotes while NHEJ contributed more in higher eukaryotes [85].

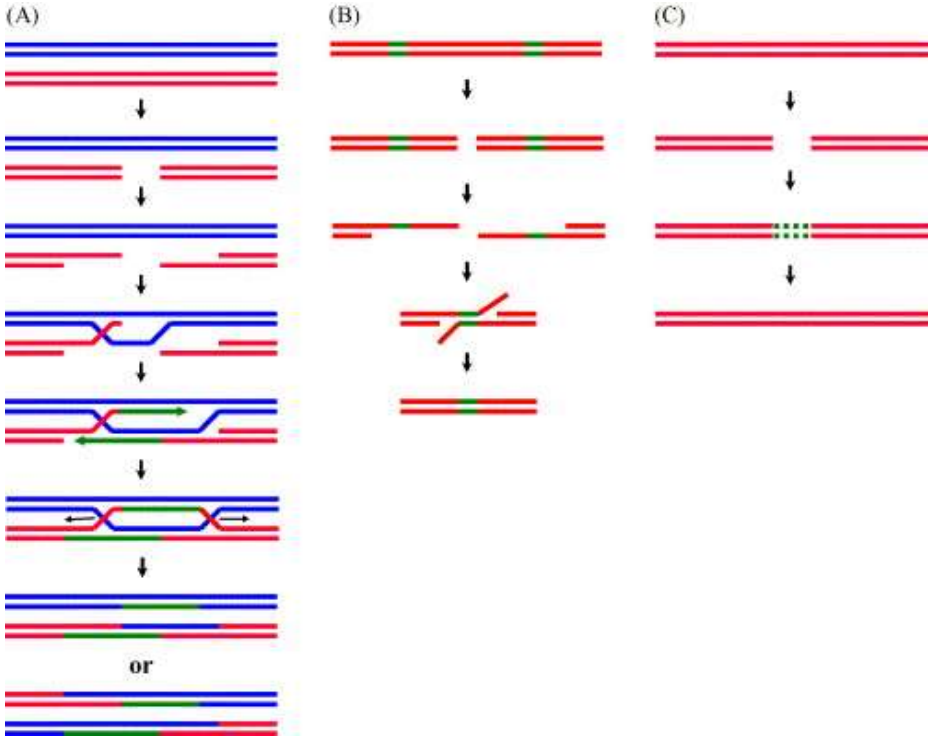


Figure 1. 5 Schemes for double DNA break repair. A. Homologous recombination (HR); B. Single-strand annealing (SSA); C. Nonhomologous DNA end joining (NHEJ). The picture was reused from [84] with permission.

1. Homologous recombination (HR)

Homologous recombination (HR) needs an intact homologous duplex as a template. Mechanically, HR is usually initiated by the enzymatic-induced degrading of 5' ends of both strands at DSB site, leaving two 3' single stranded DNA tails. Then one of the 3' ssDNA tails invades the intact homologous duplex and generates a D-loop structure, the other one anneals with the displaced strand at the joint. Using the 3'ends as a primer and the intact homologous DNA as a template, the lacking DNA fragment is

sealed by DNA replication. The resulting holliday junctions were cut by a resolvase that created a crossover or non-crossover error-free DNA repair product [84]. Deficiency of HR may result in Fanconi anemia and several types of cancers in human [86].

2. Single-strand annealing (SSA)

Single-strand annealing (SSA) doesn't need an intact homologous duplex as a template. In the processing of HR, if there are two repeated sequences existing in the 3' ssDNA tail, those two identical sequences will anneal and form a double stranded DNA fragment. Later, nonannealled ssDNA tails will be degraded away and the resulting DNA gap will be fixed through DNA replication and ligation [84]. Since one of two repeats and the sequence between the repeats are deleted, SSA is considered as an error-prone DSBR pathway [87].

3. Nonhomologous DNA end joining (NHEJ)

Without homologous DNA as template, nonhomologous DNA end joining (NHEJ) use some short ssDNA as microhomologies to guide DNA synthesis. Depending on how perfectly the microhomologies match with the damaged DNA, the repair accuracy will be different. Most of the time, NHEJ repairs the DSBs correctly. However, the noncompatability of microhomologies often lead to deletions, duplications, nucleotide substitutions and translocations. Therefore, NHEJ is also an error-prone DSBR pathway [88]. Defective NHEJ is associated with several human diseases including radiosensitive severe combined immunodeficiency and LIG4 syndrome [89].

G. SOS response

As a global response to DNA damage, SOS is triggered by the DNA polymerase stalling in the replication fork. Once large amount of ssDNA or DSBs are present in cells, the single stranded binding protein RecA will assemble into filaments and promote the self-degrading of LexA that normally binds to the promoter of SOS genes as a repressor. SOS system contains lots of DNA repair related genes including uvrABC, recA, and translesion DNA polymerases (dinA, dinB and umuDC). The uvrABC genes in the NER pathway are the first activated to defend the damages [90]. If NER pathway is not enough to fix the lesions, a full-fledge SOS will be induced by activating more SOS DNA repair genes. The error tolerant translesion DNA polymerases are able to push the replication machine forward, however, some mutagenesis might be introduced into the genome [91]. The bacteria antibiotic resistance is the result of translesion DNA polymerase-induced mutations [92].

IV. DNA deamination and uracil DNA glycosylase superfamily

A. Deamination of DNA bases

DNA base deamination is usually introduced by reactive nitrogen species and deaminase. Reactive nitrogen species (RNS), such as nitric oxide, nitrate (NO_3^-), nitrite (NO_2^-) and peroxynitrite ($ONOO^-$), are natural metabolic products *in vivo* [93]. The regular DNA bases, including adenine (A), guanine (G) and cytosine (C), are continually attacked by RNS and subjected to switch to hypoxanthine (H) (inosine (I)), xanthine (X) or oxanine (O) and uracil (U) (Fig. 1.6). Besides RNS, the widely distributed biological deaminases also contribute to DNA deamination. For instance, activation-induced cytidine deaminase (AID) turns cytosine (C) to uracil (U); Adenosine deaminase (ADA)

converts adenosine to hypoxanthine and guanine deaminase (GDA) converts guanine (G) to xanthine (X). The resulting deaminated DNA bases are potential sources of DNA transition mutations. Uracil could be mis-incorporated into A.U base pairs during DNA replication as DNA polymerase is not good at distinguishing dU and dT. The resulting A.U base pair may lead to an A.T to G.C mutation in the next DNA replication. Also, due to the deamination of cytosine in DNA strands, a G.U base pair will be generated that might result in a G.C to A.T mutation. Similar transition mutations may be caused by the other three deaminated DNA bases.

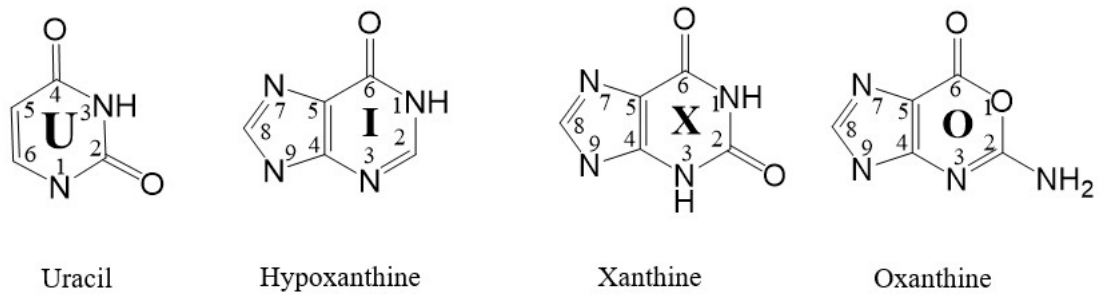


Figure 1. 6 Chemical structure of 4 deaminated DNA bases. U, uracil; I, hypoxanthine; X, xanthine; O, oxanine.

B. Uracil DNA glycosylases

To repair the deaminated DNA bases, uracil DNA glycosylase (UDG) in the base excision repair (BER) pathway plays a very crucial role. UDG is the initial enzyme in the BER pathway and is able to recognize the deaminated bases and excise them by cleaving the N-glycosidic bond. Next, Apurinic/apyrimidinic (AP) endonuclease will clear the leaving AP site. Finally, DNA polymerase and ligase will fill the gap. So far, based on

the sequence similarity of 3 functional motifs, 6 families have been identified in the UDG superfamily (Fig. 1.7). Each family presents distinct functions.

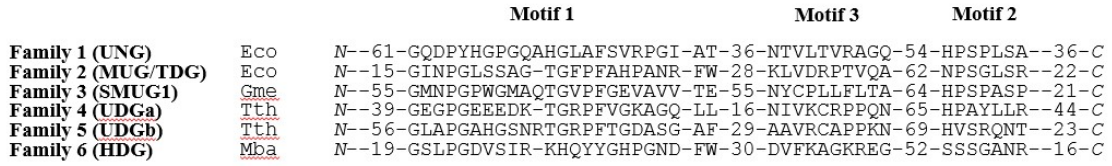


Figure 1. 7 Sequence alignment of the functional motifs from UDG superfamily. Family 1 (UDG): Eco, *E. coli*, NP_289138. Family 2 (MUG/TDG): Eco, *E. coli*, P0A9H1. Family 3 (SMUG1): Gme, *G. metallireducens* GS-15, YP_383069. Family 4 (UDGa): Tth, *T. thermophilus* HB27, YP_004341.1. Family 5 (UDGb): Tth, *T. thermophilus* HB8, YP_144415.1. Family 6 (HDG): Mba, *Methanosarcina barkeri str*. Fusaro, YP_304295.1.

1. Family 1 UNG

Family 1 UDG, termed as uracil N-glycosylase (UNG), is the first identified enzyme in the UDG superfamily [94]. UNG is ubiquitous in prokaryotes and eukaryotes. As unique structural characteristics, UNG shows extremely robust and exclusive UDG activity to both single-stranded and double-stranded uracil-containing DNA [95]. The UDG activity of UNG can be inhibited by uracil DNA glycosylase inhibitors (Ugi) from *Bacillus subtilis* bacteriophages PBS1/PBS2 in 1:1 molar stoichiometry [96]. Recently, two new inhibitors, P56 and SSP0047 were identified from *Bacillus* phages φ 29 and *Staphylococcus aureus*, respectively [97-100].

Actually, UNG not only participates in BER pathway, evidences showed UNGs are also involved in other mechanisms. In immunity, initiated by AID-induced uracil in DNA, class switch recombination (CSR) and somatic hypermutation (SHM) are two major mechanisms for the diversity of antibody. As the uracil cleavage activity, UNG is

proposed to have a regulatory role to those two mechanisms. Interestingly, in UNG deficient mice, CSR is suppressed while SHM is increased [101].

In vaccinia virus, UNG is a necessary component of the DNA replication machine and its replication processing role is independent from its UDG activity [102-104]. UNGs are also reported to be involved in DNA replication in Epstein-Barr virus and HIV-1 [105, 106]. A few newly reports about UNG might imply some other potential roles *in vivo*. For example, under treatment of pemetrexed, the increasing DNA double strand breaks (DSBs) in UNG deficient lung cancer cells implicated the repair roles of UNG to DSBs [107]. In zebrafish, knockdown of UNG increased DNA methylation level and embryonic lethality indicated UNG is indispensable for the development of zebrafish embryo [108].

2. Family 2 TDG/MUG

Family 2 UDGs are represented by *E. coli* mismatch-specific uracil DNA glycosylase (*E. coli* MUG) and human thymine DNA glycosylase (hTDG) that's why they are termed as TDG/MUG Family. Family 2 enzymes are present in bacteria, yeast, vertebrates and human. Lacking the catalytic Asp in motif 1 and His in motif 2 as in UNG, family 2 enzymes are not as robust as Family 1 UNG [108]. However, Family 2 enzymes are well known for the broad substrates. As far as we know, *E. coli* MUG is not only working as a UDG on mismatched T/U, G/U and C/U base pairs but also as XDG on all xanthine-containing DNA substrates [109]. Human TDG was initially found to cleave T from G. T mismatched base pair [110]. Later, it was reported to act toward G.U, A.U and G.I mismatched base pairs [111, 112]. Recently, the excision activity of 5-

formylcytosine and 5-caboxylcytosine (5-caC) revealed more physiological roles of hTDG in DNA demethylation pathway. Another Family 2 member *S. pombe* (Spo) TDG is able to work on all deaminated DNA bases in the order of xanthine, inosine, uracil and oxantine [112]. As matter of fact, besides deaminated DNA bases, *E. coli* MUG and hTDG were reported to work on oxidation DNA bases (such as 7,8-Dihydro-8-oxoadenine (8oxoA) and ε-adducts (such as 3,N(4)-ethenocytosine and 1,N(2)-ethenoguanine) [113-115]. Such broader activities *in vitro* indicated the crucial DNA repair roles of Family 2 MUG/TDGss *in vivo*. Besides, mammal TDGs are involved in the development of embryo. TDG deficient mice suffered severe embryonic lethality although the mutation rate of TDG (-/-) mice didn't increased significantly [116].

3. Family 3 SMUG1

Family 3 enzymes were named as single-stranded selective monofunctional UDG (SMUG1) and they are present in vertebrates, insects and some eubacteria. SMUG1 enzymes were firstly isolated from *Xenopus laevis* and human by genome-wide screening. Later, a genome database searching identified SMUG1 orthologs in a few eubacteria lacking Family 1 UNG [117]. In comparison to the robust Family 1 UNG, SMUG1 enzymes have much lower efficiency toward uracil due to higher product binding affinity, but they are capable of excising broader substrates, including xanthine, 5-formyluracil, 5-hydroxyuracil, 5-hydroxymethyluracil and 3, N^4-ethenocytosine [118-120]. Depending on reaction conditions, SMUG1 presented different substrate preference. Notably, in the presence of Mg^{2+}, SMUG1 switched into double-stranded-selective UDG [121, 122]. Revealed by depleting *Smug1* and *Smug1/Ung* from mice,

SMUG1 was reported to be the principal hmU glycosylase in mice and the major backup for UNG to excise uracil. Later on, we also reported bacterial SMUG1 is the major DNA repair enzyme for deaminated pyrimidine DNA due to the lack of UNGs in the genome of SMUG1-containing species [117].

Recently, phylogenetic analysis of UDG superfamily indicated a new group of UDG proteins, present in several eubacteria including *Listeria* genus, few species of *Lactobacillus* genus, *Streptomyces sp.*, *Amycolatopsis sp.* and *Flavobacteriaceae bacterium* [123]. Since the newly found protein group shares the same ancestral gene with family 1 and family 3 UDGs and more similar to canonical family 3 UDGs than other UDG families, it was considered as a new clade of family 3 and noted by SMUG1-like. In this thesis, the enzyme properties of a preventive of SMUG1-like enzyme, *Listeria inococa* SMUG1-like, will be discussed in chapter 2.

4. Family 4 UDGa

Like Family 1 UNG, Family 4 UDGa only works on uracil-contain DNA regardless double-stranded or single-stranded. The first Family 4 UDG enzyme was identified from the extreme thermophile *Thermotoga maritima* in which a uracil-removing activity was present, however, no family 1 homolog was found in the genome [124]. Later, more Family 4 members are isolated from some thermophilic archaea and bacteria, such as *Archaeoglobus fulgidus, Pyrobaculum aerophilum* and *Thermus thermophilus* [125-127]. All those thermophilic organisms are at very high risk of cytosine deamination due to the higher growing temperature [128]. Family 4 UDGa showed a different degree of thermostability: *T. martitima* UDG is even active after pre-

incubation at 75 ℃ and *A. fulgidus* UDG still cleaves uracil at 50 ℃ [124, 129]. Structurally, Family 4 UDGa contains a [4Fe-4S] cluster which is not crucial for the UDG activity but is expected to be important to stabilize the protein conformation. The exact functions of [4Fe-4S] cluster are still not sure. By the comparison of substrate recognition of family 4 *T. thermophilus* UDG and other UDG families, the catalytic mechanisms of family 4 UDG are very similar to family 1 UNG as shown in Fig. 1.8 [127]. Both of them contain a Phe, His, Asn and Gln interacting with uracil in a similar way. Meanwhile, Glu 47 plays an analogical function with Try 147, both of them block the substrate recognition pocket that explained why family 4 UDGa only works on uracil as family 1 UNG. Recently, site-directed mutations, mutual information and molecular dynamics analysis studies indicated that the highly conserved residues E and G in the motif 1 of family 4 UDGa are correlated [130].

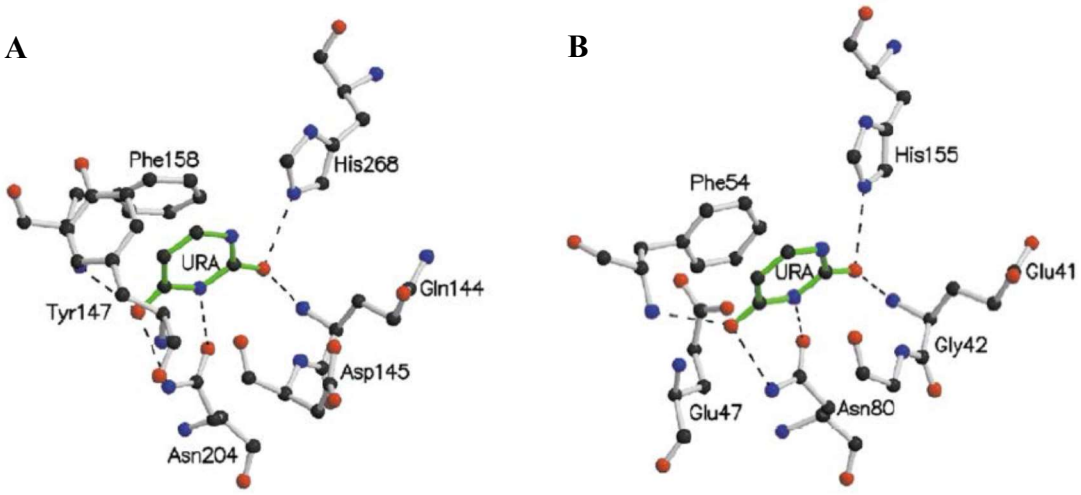


Figure 1. 8 Active sites of Family 1 UNG and Family 4 UDGa. The active site of Family 1 UNG with uracil from human; B. The active site of Family 4 UDGa with uracil from *Thermus thermophilus*. Picture is reused from [127] with permission.

5. Family 5 UDGb

Family 5 UDGb enzymes are also mainly found in thermophilic organisms. However, unlike the uracil-specific Family 4 UDGa, Family 5 UDGb enzymes are versatile. For instance, *P. aerophilum* UDGb cleaves uracil, hydroxymethyluracil, fluorouracil and hypoxanthine from DNA; *Mycobacterium tuberculosis* UDGb cleaves uracil, ethenocytosine and hypoxanthine from DNA; *T. thermophilus* UDGb cleaves uracil, hypoxanthine and xanthine from DNA [131-133]. Such broader substrate specificity implicit the essential DNA repair functions of UDGb *in vivo*. UDGb deficient *Mycobacterium smegmatis* shows higher spontaneous mutation rate resulting from the deamination of cytosine and adenine [134]. Disruption of UDGb from *Thermus thermophilus* result in 3-folds increase of mutation rate [135]. Recently, a genetic assay revealed that UDGb is able to remove the misincorporated uracil from A.U base pairs *in vivo* [132]. Structurally, lacking Asp/Asn as the catalytic residue in motif 1, Asn120 in Tth UDGb at motif 3 was identified as a residue that activate a water molecular and cleave the N-glycosidic bond [132].

6. Family 6 HDG [136]

All of the above discussed UDG families have UDG activity, however, the newly found family 6 UDG, widely exist in archaea, eubacteria and eukaryotes, doesn't work as UDG but a hypoxanthine DNA glycosylase (HDG). The first family 6 HDG was isolated from *M. barkeri* that mainly works on double-stranded hypoxanthine-containing DNA substrates in the order of G/I > T/I > A/I > C/I. Meanwhile, a *lac*-Z based genetic assay confirmed that the HDG activity of *M. barkeri* HDG *in vivo*. Structurally, family 6 UDG

contains a distinct motif 1 and motif 2. Notably, family 6 UDGs have no Asp/Asn in motif 1 or His in motif 2 as other families. Mutagenesis and molecular modelling work proposed the highly conserved Asn 39 in *M. barkeri* HDG as the catalytic residue although this might need further confirmation.

V. References

1. Devasagayam, T. P., Tilak, J. C., Boloor, K. K., Sane, K. S., Ghaskadbi, S. S. & Lele, R. D. (2004) Free radicals and antioxidants in human health: current status and future prospects, *The Journal of the Association of Physicians of India.* **52**, 794-804.

2. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. (2003) Oxidative DNA damage: mechanisms, mutation, and disease, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology.* **17**, 1195-214.

3. Helbock, H. J., Beckman, K. B. & Ames, B. N. (1999) 8-Hydroxydeoxyguanosine and 8-hydroxyguanine as biomarkers of oxidative DNA damage, *Methods in enzymology.* **300**, 156-66.

4. Dizdaroglu, M. (1992) Oxidative damage to DNA in mammalian chromatin, *Mutation research.* **275**, 331-42.

5. Kuchino, Y., Mori, F., Kasai, H., Inoue, H., Iwai, S., Miura, K., Ohtsuka, E. & Nishimura, S. (1987) Misreading of DNA templates containing 8-hydroxydeoxyguanosine at the modified base and at adjacent residues, *Nature.* **327**, 77-9.

6. Cheng, K. C., Cahill, D. S., Kasai, H., Nishimura, S. & Loeb, L. A. (1992) 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions, *The Journal of biological chemistry.* **267**, 166-72.

7. Evans, M. D., Dizdaroglu, M. & Cooke, M. S. (2004) Oxidative DNA damage and disease: induction, repair and significance, *Mutation research.* **567**, 1-61.

8. Lyras, L., Perry, R. H., Perry, E. K., Ince, P. G., Jenner, A., Jenner, P. & Halliwell, B. (1998) Oxidative damage to proteins, lipids, and DNA in cortical brain regions from patients with dementia with Lewy bodies, *Journal of neurochemistry.* **71**, 302-12.

9. Rehman, A., Nourooz-Zadeh, J., Moller, W., Tritschler, H., Pereira, P. & Halliwell, B. (1999) Increased oxidative damage to all DNA bases in patients with type II diabetes mellitus, *FEBS letters.* **448**, 120-2.

10. Frenkel, K., Karkoszka, J., Kim, E. & Taioli, E. (1993) Recognition of oxidized DNA bases by sera of patients with inflammatory diseases, *Free radical biology & medicine.* **14**, 483-94.

11. Rydberg, B. & Lindahl, T. (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction, *The EMBO journal.* **1**, 211-6.

12. Barrows, L. R. & Magee, P. N. (1982) Nonenzymatic methylation of DNA by S-adenosylmethionine in vitro, *Carcinogenesis.* **3**, 349-51.

13. Drablos, F., Feyzi, E., Aas, P. A., Vaagbo, C. B., Kavli, B., Bratlie, M. S., Pena-Diaz, J., Otterlei, M., Slupphaug, G. & Krokan, H. E. (2004) Alkylation damage in DNA and RNA--repair mechanisms and medical significance, *DNA repair.* **3**, 1389-407.

14. Hecht, S. S. (1999) DNA adduct formation from tobacco-specific N-nitrosamines, *Mutation research.* **424**, 127-42.

15. Lindahl, T. & Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid, *Biochemistry.* **11**, 3610-8.

16. Scharer, O. D. & Jiricny, J. (2001) Recent progress in the biology, chemistry and structural biology of DNA glycosylases, *BioEssays : news and reviews in molecular, cellular and developmental biology.* **23**, 270-81.

17. Yu, S. L., Lee, S. K., Johnson, R. E., Prakash, L. & Prakash, S. (2003) The stalling of transcription at abasic sites is highly mutagenic, *Molecular and cellular biology.* **23**, 382-8.

18. Prakash, S. & Prakash, L. (2002) Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair, *Genes & development.* **16**, 1872-83.

19. Caldecott, K. W. (2001) Mammalian DNA single-strand break repair: an X-ra(y)ted affair, *BioEssays : news and reviews in molecular, cellular and developmental biology.* **23**, 447-55.

20. Sinha, R. P. & Hader, D. P. (2002) UV-induced DNA damage and repair: a review, *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology.* **1**, 225-36.

21. Lee, J. H., Hwang, G. S. & Choi, B. S. (1999) Solution structure of a DNA decamer duplex containing the stable 3' T.G base pair of the pyrimidine(6-4)pyrimidone photoproduct [(6-4) adduct]: implications for the highly specific 3' T --> C transition of the (6-4) adduct, *Proceedings of the National Academy of Sciences of the United States of America.* **96**, 6632-6.

22. Lee, J. H., Bae, S. H. & Choi, B. S. (2000) The Dewar photoproduct of thymidylyl(3'-->5')- thymidine (Dewar product) exhibits mutagenic behavior in accordance with the "A rule", *Proceedings of the National Academy of Sciences of the United States of America.* **97**, 4591-6.

23. Lee, J. H., Hwang, G. S., Kim, J. K. & Choi, B. S. (1998) The solution structure of DNA decamer duplex containing the Dewar product of thymidylyl(3-->5')thymidine by NMR and full relaxation matrix refinement, *FEBS letters.* **428**, 269-74.

24. Banath, J. P., Banuelos, C. A., Klokov, D., MacPhail, S. M., Lansdorp, P. M. & Olive, P. L. (2009) Explanation for excessive DNA single-strand breaks and endogenous repair foci in pluripotent mouse embryonic stem cells, *Experimental cell research.* **315**, 1505-20.

25. Demple, B. & DeMott, M. S. (2002) Dynamics and diversions in base excision DNA repair of oxidized abasic lesions, *Oncogene.* **21**, 8926-34.

26. Hegde, M. L., Hazra, T. K. & Mitra, S. (2008) Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells, *Cell research.* **18**, 27-47.

27. Wang, J. C. (2002) Cellular roles of DNA topoisomerases: a molecular perspective, *Nature reviews Molecular cell biology.* **3**, 430-40.

28. Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 8241-6.

29. Kouzminova, E. A. & Kuzminov, A. (2006) Fragmentation of replicating chromosomes triggered by uracil in DNA, *Journal of molecular biology.* **355**, 20-33.

30. Zhou, W. & Doetsch, P. W. (1994) Transcription bypass or blockage at single-strand breaks on the DNA template strand: effect of different 3' and 5' flanking groups on the T7 RNA polymerase elongation complex, *Biochemistry.* **33**, 14926-34.

31. Kathe, S. D., Shen, G. P. & Wallace, S. S. (2004) Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts, *The Journal of biological chemistry.* **279**, 18511-20.

32. Caldecott, K. W. (2008) Single-strand break repair and genetic disease, *Nature reviews Genetics.* **9**, 619-31.

33. Khanna, K. K. & Jackson, S. P. (2001) DNA double-strand breaks: signaling, repair and the cancer connection, *Nature genetics.* **27**, 247-54.

34. Lieber, M. R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway, *Annual review of biochemistry.* **79**, 181-211.

35. Lin, W. Y., Wilson, J. H. & Lin, Y. (2013) Repair of chromosomal double-strand breaks by precise ligation in human cells, *DNA repair.* **12**, 480-7.

36. Adachi, N., Suzuki, H., Iiizumi, S. & Koyama, H. (2003) Hypersensitivity of nonhomologous DNA end-joining mutants to VP-16 and ICRF-193: implications for the repair of topoisomerase II-mediated DNA damage, *The Journal of biological chemistry.* **278**, 35897-902.

37. Mahowald, G. K., Baron, J. M. & Sleckman, B. P. (2008) Collateral damage from antigen receptor gene diversification, *Cell.* **135**, 1009-12.

38. Ferguson, D. O. & Alt, F. W. (2001) DNA double strand break repair and chromosomal translocation: lessons from animal models, *Oncogene.* **20**, 5572-9.

39. Bohgaki, T., Bohgaki, M. & Hakem, R. (2010) DNA double-strand break signaling and human disorders, *Genome integrity.* **1**, 15.

40. Liu, Y., Prasad, R., Beard, W. A., Kedar, P. S., Hou, E. W., Shock, D. D. & Wilson, S. H. (2007) Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta, *The Journal of biological chemistry.* **282**, 13532-41.

41. Lu, R., Nash, H. M. & Verdine, G. L. (1997) A mammalian DNA repair enzyme that excises oxidatively damaged guanines maps to a locus frequently lost in lung cancer, *Current biology : CB.* **7**, 397-407.

42. van der Kemp, P. A., Thomas, D., Barbey, R., de Oliveira, R. & Boiteux, S. (1996) Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine, *Proceedings of the National Academy of Sciences of the United States of America.* **93**, 5197-202.

43. Dherin, C., Radicella, J. P., Dizdaroglu, M. & Boiteux, S. (1999) Excision of oxidatively damaged DNA bases by the human alpha-hOgg1 protein and the polymorphic alpha-hOgg1(Ser326Cys) protein which is frequently found in human populations, *Nucleic acids research.* **27**, 4001-7.

44. Karahalil, B., Girard, P. M., Boiteux, S. & Dizdaroglu, M. (1998) Substrate specificity of the Ogg1 protein of Saccharomyces cerevisiae: excision of guanine lesions produced in DNA by ionizing radiation- or hydrogen peroxide/metal ion-generated free radicals, *Nucleic acids research.* **26**, 1228-33.

45. Sampath, H., Vartanian, V., Rollins, M. R., Sakumi, K., Nakabeppu, Y. & Lloyd, R. S. (2012) 8-Oxoguanine DNA glycosylase (OGG1) deficiency increases susceptibility to obesity and metabolic dysfunction, *PloS one.* **7**, e51697.

46. Klungland, A., Rosewell, I., Hollenbach, S., Larsen, E., Daly, G., Epe, B., Seeberg, E., Lindahl, T. & Barnes, D. E. (1999) Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage, *Proceedings of the National Academy of Sciences of the United States of America.* **96**, 13300-5.

47. Janik, J., Swoboda, M., Janowska, B., Ciesla, J. M., Gackowski, D., Kowalewski, J., Olinski, R., Tudek, B. & Speina, E. (2011) 8-Oxoguanine incision activity is impaired in lung tissues of NSCLC patients with the polymorphism of OGG1 and XRCC1 genes, *Mutation research.* **709-710**, 21-31.

48. Stanczyk, M., Sliwinski, T., Cuchra, M., Zubowska, M., Bielecka-Kowalska, A., Kowalski, M., Szemraj, J., Mlynarski, W. & Majsterek, I. (2011) The association of polymorphisms in DNA base excision repair genes XRCC1, OGG1 and MUTYH with the risk of childhood acute lymphoblastic leukemia, *Molecular biology reports.* **38**, 445-51.

49. McGoldrick, J. P., Yeh, Y. C., Solomon, M., Essigmann, J. M. & Lu, A. L. (1995) Characterization of a mammalian homolog of the Escherichia coli MutY mismatch repair protein, *Molecular and cellular biology.* **15**, 989-96.

50. Au, K. G., Cabrera, M., Miller, J. H. & Modrich, P. (1988) Escherichia coli mutY gene product is required for specific A-G----C.G mismatch correction, *Proceedings of the National Academy of Sciences of the United States of America.* **85**, 9163-6.

51. Markkanen, E., Dorn, J. & Hubscher, U. (2013) MUTYH DNA glycosylase: the rationale for removing undamaged bases from the DNA, *Frontiers in genetics.* **4**, 18.

52. Saparbaev, M. & Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases, *Proceedings of the National Academy of Sciences of the United States of America.* **91**, 5873-7.

53. Bessho, T., Roy, R., Yamamoto, K., Kasai, H., Nishimura, S., Tano, K. & Mitra, S. (1993) Repair of 8-hydroxyguanine in DNA by mammalian N-methylpurine-DNA glycosylase, *Proceedings of the National Academy of Sciences of the United States of America.* **90**, 8901-4.

54. Bjelland, S., Birkeland, N. K., Benneche, T., Volden, G. & Seeberg, E. (1994) DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in Escherichia coli, *The Journal of biological chemistry.* **269**, 30489-95.

55. Helland, D. E., Male, R., Haukanes, B. I., Olsen, L., Haugan, I. & Kleppe, K. (1987) Properties and mechanism of action of eukaryotic 3-methyladenine-DNA glycosylases, *Journal of cell science Supplement.* **6**, 139-46.

56. Grzesiuk, E., Gozdek, A. & Tudek, B. (2001) Contribution of E. coli AlkA, TagA glycosylases and UvrABC-excinuclease in MMS mutagenesis, *Mutation research.* **480-481**, 77-84.

57. Hendricks, C. A., Razlog, M., Matsuguchi, T., Goyal, A., Brock, A. L. & Engelward, B. P. (2002) The S. cerevisiae Mag1 3-methyladenine DNA glycosylase modulates susceptibility to homologous recombination, *DNA repair.* **1**, 645-59.

58. Daley, J. M., Zakaria, C. & Ramotar, D. (2010) The endonuclease IV family of apurinic/apyrimidinic endonucleases, *Mutation research.* **705**, 217-27.

59. Ramotar, D., Popoff, S. C., Gralla, E. B. & Demple, B. (1991) Cellular role of yeast Apn1 apurinic endonuclease/3'-diesterase: repair of oxidative and alkylation DNA damage and control of spontaneous mutation, *Molecular and cellular biology.* **11**, 4537-44.

60. Zakaria, C., Kassahun, H., Yang, X., Labbe, J. C., Nilsen, H. & Ramotar, D. (2010) Caenorhabditis elegans APN-1 plays a vital role in maintaining genome stability, *DNA repair.* **9**, 169-76.

61. Ramotar, D., Vadnais, J., Masson, J. Y. & Tremblay, S. (1998) Schizosaccharomyces pombe apn1 encodes a homologue of the Escherichia coli endonuclease IV family of DNA repair proteins, *Biochimica et biophysica acta.* **1396**, 15-20.

62. Mol, C. D., Hosfield, D. J. & Tainer, J. A. (2000) Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means, *Mutation research.* **460**, 211-29.

63. Morris, L. P., Degtyareva, N., Sheppard, C., Heyburn, L., Ivanov, A. A., Kow, Y. W. & Doetsch, P. W. (2012) Saccharomyces cerevisiae Apn1 mutation affecting stable protein expression mimics catalytic activity impairment: implications for assessing DNA repair capacity in humans, *DNA repair.* **11**, 753-65.

64. Park, J. S., Kim, H. L., Kim, Y. J., Weon, J. I., Sung, M. K., Chung, H. W. & Seo, Y. R. (2014) Human AP endonuclease 1: a potential marker for the prediction of environmental carcinogenesis risk, *Oxidative medicine and cellular longevity.* **2014**, 730301.

65. Suzuki, T., Yamamoto, K., Harashima, H. & Kamiya, H. (2008) Base excision repair enzyme endonuclease III suppresses mutagenesis caused by 8-hydroxy-dGTP, *DNA repair.* **7**, 88-94.

66. Moolla, N., Goosens, V. J., Kana, B. D. & Gordhan, B. G. (2014) The contribution of Nth and Nei DNA glycosylases to mutagenesis in Mycobacterium smegmatis, *DNA repair.* **13**, 32-41.

67. Hazra, T. K. & Mitra, S. (2006) Purification and characterization of NEIL1 and NEIL2, members of a distinct family of mammalian DNA glycosylases for repair of oxidized bases, *Methods in enzymology.* **408**, 33-48.

68. Matsumoto, Y., Zhang, Q. M., Takao, M., Yasui, A. & Yonei, S. (2001) Escherichia coli Nth and human hNTH1 DNA glycosylases are involved in removal of 8-oxoguanine from 8-oxoguanine/guanine mispairs in DNA, *Nucleic acids research.* **29**, 1975-81.

69. Broderick, P., Bagratuni, T., Vijayakrishnan, J., Lubbe, S., Chandler, I. & Houlston, R. S. (2006) Evaluation of NTHL1, NEIL1, NEIL2, MPG, TDG, UNG and SMUG1 genes in familial colorectal cancer predisposition, *BMC cancer.* **6**, 243.

70. Weren, R. D., Ligtenberg, M. J., Kets, C. M., de Voer, R. M., Verwiel, E. T., Spruijt, L., van Zelst-Stams, W. A., Jongmans, M. C. & Gilissen, C. (2015) A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. **47**, 668-71.

71. Essen, L. O. & Klar, T. (2006) Light-driven DNA repair by photolyases, *Cellular and molecular life sciences : CMLS.* **63**, 1266-77.

72. Johnson, J. L., Hamm-Alvarez, S., Payne, G., Sancar, G. B., Rajagopalan, K. V. & Sancar, A. (1988) Identification of the second chromophore of Escherichia coli and yeast DNA photolyases as 5,10-methenyltetrahydrofolate, *Proceedings of the National Academy of Sciences of the United States of America.* **85**, 2046-50.

73. Eker, A. P., Kooiman, P., Hessels, J. K. & Yasui, A. (1990) DNA photoreactivating enzyme from the cyanobacterium Anacystis nidulans, *The Journal of biological chemistry.* **265**, 8009-15.

74. Ueda, T., Kato, A., Kuramitsu, S., Terasawa, H. & Shimada, I. (2005) Identification and characterization of a second chromophore of DNA photolyase from Thermus thermophilus HB27, *The Journal of biological chemistry.* **280**, 36237-43.

75. Saxena, C., Sancar, A. & Zhong, D. P. (2004) Femtosecond dynamics of DNA photolyase: Energy transfer of antenna initiation and electron transfer of cofactor reduction, *J Phys Chem B.* **108**, 18026-18033.

76. Kisker, C., Kuper, J. & Van Houten, B. (2013) Prokaryotic nucleotide excision repair, *Cold Spring Harbor perspectives in biology.* **5**, a012591.

77. Maddukuri, L., Dudzinska, D. & Tudek, B. (2007) Bacterial DNA repair genes and their eukaryotic homologues: 4. The role of nucleotide excision DNA repair (NER) system in mammalian cells, *Acta biochimica Polonica.* **54**, 469-82.

78. Erie, D. A. & Weninger, K. R. (2014) Single molecule studies of DNA mismatch repair, *DNA repair.* **20**, 71-81.

79. Jiricny, J. (2006) The multifaceted mismatch-repair system, *Nature reviews Molecular cell biology.* **7**, 335-46.

80. Peltomaki, P. & Vasen, H. (2004) Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database, *Disease markers.* **20**, 269-76.

81. Caldecott, K. W. (2003) XRCC1 and DNA strand break repair, *DNA repair.* **2**, 955-69.

82. Date, H., Onodera, O., Tanaka, H., Iwabuchi, K., Uekawa, K., Igarashi, S., Koike, R., Hiroi, T., Yuasa, T., Awaya, Y., Sakai, T., Takahashi, T., Nagatomo, H., Sekijima, Y., Kawachi, I., Takiyama, Y., Nishizawa, M., Fukuhara, N., Saito, K., Sugano, S. & Tsuji, S. (2001) Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene, *Nature genetics.* **29**, 184-8.

83. Interthal, H., Chen, H. J., Kehl-Fie, T. E., Zotzmann, J., Leppard, J. B. & Champoux, J. J. (2005) SCAN1 mutant Tdp1 accumulates the enzyme--DNA intermediate and causes camptothecin hypersensitivity, *The EMBO journal.* **24**, 2224-33.

84. Dudas, A. & Chovanec, M. (2004) DNA double-strand break repair by homologous recombination, *Mutation research.* **566**, 131-67.

85. Critchlow, S. E. & Jackson, S. P. (1998) DNA end-joining: from yeast to man, *Trends in biochemical sciences.* **23**, 394-8.

86. Phillips, E. R. & McKinnon, P. J. (2007) DNA double-strand break repair and development, *Oncogene.* **26**, 7799-808.

87. Helleday, T., Lo, J., van Gent, D. C. & Engelward, B. P. (2007) DNA double-strand break repair: from mechanistic understanding to cancer treatment, *DNA repair.* **6**, 923-35.

88. Boulton, S. J. & Jackson, S. P. (1996) Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways, *The EMBO journal.* **15**, 5093-103.

89. Kerzendorfer, C. & O'Driscoll, M. (2009) Human DNA damage response and repair deficiency syndromes: linking genomic instability and cell cycle checkpoint proficiency, *DNA repair.* **8**, 1139-52.

90. Michel, B. (2005) After 30 years of study, the bacterial SOS response still surprises us, *PLoS biology.* **3**, e255.

91. Andersson, D. I. & Hughes, D. (2014) Microbiological effects of sublethal levels of antibiotics, *Nature reviews Microbiology.* **12**, 465-78.

92. Lee, A. M., Ross, C. T., Zeng, B. B. & Singleton, S. F. (2005) A molecular target for suppression of the evolution of antibiotic resistance: inhibition of the Escherichia coli RecA protein by N(6)-(1-naphthyl)-ADP, *Journal of medicinal chemistry.* **48**, 5408-11.

93. Squadrito, G. L. & Pryor, W. A. (1998) Oxidative chemistry of nitric oxide: the roles of superoxide, peroxynitrite, and carbon dioxide, *Free radical biology & medicine.* **25**, 392-403.

94. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues, *Proceedings of the National Academy of Sciences of the United States of America.* **71**, 3649-53.

95. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase, *Nature.* **373**, 487-93.

96. Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA, *Cell.* **82**, 701-8.

97. Wang, H. C., Hsu, K. C., Yang, J. M., Wu, M. L., Ko, T. P., Lin, S. R. & Wang, A. H. (2014) Staphylococcus aureus protein SAUGI acts as a uracil-DNA glycosylase inhibitor, *Nucleic acids research.* **42**, 1354-64.

98. Serrano-Heras, G., Ruiz-Maso, J. A., del Solar, G., Espinosa, M., Bravo, A. & Salas, M. (2007) Protein p56 from the Bacillus subtilis phage phi29 inhibits DNA-binding ability of uracil-DNA glycosylase, *Nucleic acids research.* **35**, 5393-401.

99. Perez-Lago, L., Serrano-Heras, G., Banos, B., Lazaro, J. M., Alcorlo, M., Villar, L. & Salas, M. (2011) Characterization of Bacillus subtilis uracil-DNA glycosylase and its inhibition by phage phi29 protein p56, *Mol Microbiol.* **80**, 1657-66.

100. Banos-Sanz, J. I., Mojardin, L., Sanz-Aparicio, J., Lazaro, J. M., Villar, L., Serrano-Heras, G., Gonzalez, B. & Salas, M. (2013) Crystal structure and functional insights into uracil-DNA glycosylase inhibition by phage Phi29 DNA mimic protein p56, *Nucleic acids research.* **41**, 6761-73.

101. Yousif, A. S., Stanlie, A., Begum, N. A. & Honjo, T. (2014) Opinion: uracil DNA glycosylase (UNG) plays distinct and non-canonical roles in somatic hypermutation and class switch recombination, *International immunology.* **26**, 575-8.

102. Boyle, K. A., Stanitsa, E. S., Greseth, M. D., Lindgren, J. K. & Traktman, P. (2011) Evaluation of the role of the vaccinia virus uracil DNA glycosylase and A20 proteins as intrinsic components of the DNA polymerase holoenzyme, *The Journal of biological chemistry.* **286**, 24702-13.

103. De Silva, F. S. & Moss, B. (2003) Vaccinia virus uracil DNA glycosylase has an essential role in DNA synthesis that is independent of its glycosylase activity: catalytic site mutations reduce virulence but not virus replication in cultured cells, *Journal of virology.* **77**, 159-66.

104. Stanitsa, E. S., Arps, L. & Traktman, P. (2006) Vaccinia virus uracil DNA glycosylase interacts with the A20 protein to form a heterodimeric processivity factor for the viral DNA polymerase, *The Journal of biological chemistry.* **281**, 3439-51.

105. Su, M. T., Liu, I. H., Wu, C. W., Chang, S. M., Tsai, C. H., Yang, P. W., Chuang, Y. C., Lee, C. P. & Chen, M. R. (2014) Uracil DNA glycosylase BKRF3 contributes to Epstein-Barr virus DNA replication through physical interactions with proteins in viral DNA replication complex, *Journal of virology.* **88**, 8883-99.

106. Guenzel, C. A., Herate, C., Le Rouzic, E., Maidou-Peindara, P., Sadler, H. A., Rouyez, M. C., Mansky, L. M. & Benichou, S. (2012) Recruitment of the nuclear form of uracil DNA glycosylase into virus particles participates in the full infectivity of HIV-1, *Journal of virology.* **86**, 2533-44.

107. Weeks, L. D., Zentner, G. E., Scacheri, P. C. & Gerson, S. L. (2014) Uracil DNA glycosylase (UNG) loss enhances DNA double strand break formation in human cancer cells exposed to pemetrexed, *Cell death & disease.* **5**, e1045.

108. Wu, D., Chen, L., Sun, Q., Wu, X., Jia, S. & Meng, A. (2014) Uracil-DNA glycosylase is involved in DNA demethylation and required for embryonic development in the zebrafish embryo, *The Journal of biological chemistry.* **289**, 15463-73.

109. Lee, H. W., Brice, A. R., Wright, C. B., Dominy, B. N. & Cao, W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase, *The Journal of biological chemistry.* **285**, 41483-90.

110. Brown, T. C. & Jiricny, J. (1987) A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine, *Cell.* **50**, 945-50.

111. Neddermann, P. & Jiricny, J. (1994) Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells, *Proceedings of the National Academy of Sciences of the United States of America.* **91**, 1642-6.

112. Dong, L., Mi, R., Glass, R. A., Barry, J. N. & Cao, W. (2008) Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase, *DNA repair.* **7**, 1962-72.

113. Saparbaev, M., Langouet, S., Privezentzev, C. V., Guengerich, F. P., Cai, H., Elder, R. H. & Laval, J. (2002) 1,N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of Escherichia coli mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase, *The Journal of biological chemistry.* **277**, 26987-93.

114. Saparbaev, M. & Laval, J. (1998) 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 8508-13.

115. Talhaoui, I., Couve, S., Ishchenko, A. A., Kunz, C., Schar, P. & Saparbaev, M. (2013) 7,8-Dihydro-8-oxoadenine, a highly mutagenic adduct, is repaired by Escherichia coli and human mismatch-specific uracil/thymine-DNA glycosylases, *Nucleic acids research.* **41**, 912-23.

116. Saito, Y., Ono, T., Takeda, N., Nohmi, T., Seki, M., Enomoto, T., Noda, T. & Uehara, Y. (2012) Embryonic lethality in mice lacking mismatch-specific thymine DNA glycosylase is partially prevented by DOPS, a precursor of noradrenaline, *The Tohoku journal of experimental medicine.* **226**, 75-83.

117. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N. & Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG, *Journal of molecular biology.* **385**, 761-78.

118. An, Q., Robins, P., Lindahl, T. & Barnes, D. E. (2007) 5-Fluorouracil incorporated into DNA is excised by the Smug1 DNA glycosylase to reduce drug cytotoxicity, *Cancer research.* **67**, 940-5.

119. Boorstein, R. J., Cummings, A., Jr., Marenstein, D. R., Chan, M. K., Ma, Y., Neubert, T. A., Brown, S. M. & Teebor, G. W. (2001) Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1, *The Journal of biological chemistry.* **276**, 41991-7.

120. Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A. & Ide, H. (2003) Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions, *Biochemistry.* **42**, 5003-12.

121. Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E. & Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms, *Nucleic acids research.* **35**, 3879-92.

122. Kavli, B., Andersen, S., Otterlei, M., Liabakk, N. B., Imai, K., Fischer, A., Durandy, A., Krokan, H. E. & Slupphaug, G. (2005) B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil, *The Journal of experimental medicine.* **201**, 2011-21.

123. Lucas-Lledo, J. I., Maddamsetti, R. & Lynch, M. (2011) Phylogenomic analysis of the uracil-DNA glycosylase superfamily, *Molecular biology and evolution.* **28**, 1307-17.

124. Sandigursky, M. & Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes, *Current biology : CB.* **9**, 531-4.

125. Sandigursky, M. & Franklin, W. A. (2000) Uracil-DNA glycosylase in the extreme thermophile Archaeoglobus fulgidus, *The Journal of biological chemistry.* **275**, 19146-9.

126. Sartori, A. A., Schar, P., Fitz-Gibbon, S., Miller, J. H. & Jiricny, J. (2001) Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon Pyrobaculum aerophilum, *The Journal of biological chemistry.* **276**, 29979-86.

127. Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S. & Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8, *Journal of molecular biology.* **333**, 515-26.

128. Lindahl, T. & Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid, *Biochemistry.* **13**, 3405-10.

129. Knaevelsrud, I., Ruoff, P., Anensen, H., Klungland, A., Bjelland, S. & Birkeland, N. K. (2001) Excision of uracil from DNA by the hyperthermophilic Afung protein is dependent on the opposite base and stimulated by heat-induced transition to a more open structure, *Mutation research.* **487**, 173-90.

130. Xia, B., Liu, Y., Guevara, J., Li, J., Jilich, C., Yang, Y., Wang, L., Dominy, B. N. & Cao, W. (2017) Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily, *Sci Rep.* **7**, 45978.

131. Srinath, T., Bharti, S. K. & Varshney, U. (2007) Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis, *DNA repair.* **6**, 1517-28.

132. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N. & Cao, W. (2014) Specificity and catalytic mechanism in family 5 uracil DNA glycosylase, *The Journal of biological chemistry.* **289**, 18413-26.

133. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H. & Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site, *The EMBO journal.* **21**, 3182-91.

134. Wanner, R. M., Castor, D., Guthlein, C., Bottger, E. C., Springer, B. & Jiricny, J. (2009) The uracil DNA glycosylase UdgB of Mycobacterium smegmatis protects the organism from the mutagenic effects of cytosine and adenine deamination, *Journal of bacteriology.* **191**, 6312-9.

135. Sakai, T., Tokishita, S., Mochizuki, K., Motomiya, A., Yamagata, H. & Ohta, T. (2008) Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus, *DNA repair.* **7**, 663-9.

136. Lee, H. W., Dominy, B. N. & Cao, W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily, *The Journal of biological chemistry.* **286**, 31282-7.

CHAPTER TWO

IDENTIFICATION OF A PROTOTYPICAL SINGLE-STRANDED URACIL DNA GLYCOSYLASE FROM LISTERIA INNOCUA

I. Abstract

A recent phylogenetic study of UDG superfamily estimated a new clade of family 3 enzymes (SMUG1-like), which shares a lower homology with canonic SMUG1 enzymes. The enzymatic properties of the newly found putative uracil DNA glycosylase are unknown. To test the potential UDG activity and evaluate phylogenetic classification, we isolated one SMUG1-like glycosylase representative from *Listeria innocua* (Lin). A biochemical screening of DNA glycosylase activity *in vitro* indicates that Lin SMUG1-like glycosylase is a single-strand-selective uracil DNA glycosylase. The UDG activity on DNA bubble structures provides clue to its physiological significance *in vivo*. Mutagenesis and molecular modeling analysis reveal that Lin SMUG1-like glycosylase has similar functional motifs with SMUG1 enzymes; however, it contains a distinct catalytic doublet S67-S68 in motif 1 that is not found in any families in the UDG superfamily. Experimental investigation shows that the S67M-S68N double mutant is catalytically more active than either S67M or S68N single mutant. Coupled with mutual information analysis, the results indicate a high degree of correlation in the evolution of SMUG1-like enzymes. This study underscores the functional and catalytic diversity in the evolution of enzymes in UDG superfamily.

II. Introduction

As a non-canonical base of DNA, uracil can occur in DNA by misincorporation of dUTP during DNA replication [1, 2] or spontaneous or enzymatic deamination of cytosine [3-5]. Uracil DNA glycosylase (UDG) is the principal enzyme to remove uracil in DNA by hydrolyzing the N-glycosidic bond and initiating base excision repair (BER) pathway [6, 7]. Based on the sequence homology and structural fold, six families were identified. Family 1 UDGs, also known as UNGs, are ubiquitous in bacteria and most eukaryotes. The enzymes exclusively cleave uracil in both double-stranded (ds) and single-stranded (ss) DNA at a very efficient rate [2, 8]. Family 2 UDGs are represented by human thymine DNA glycosylase (hTDG) [9] and *E. coli* mismatch-specific uracil DNA glycosylase (*E. coli* MUG) [10]. They are identified as repair enzymes acting on uracil base in a mispair, xanthine-containing DNA and other modified bases [10-15]. The discovery of human TDG as a 5-formylcytosine (fC) and 5-carboxylcytosine (5-caC) DNA glycosylase places it as an essential enzyme in DNA demethylation [16, 17]. Family 3 enzymes are named as single-strand-selective monofunctional UDG (SMUG1) and are present in vertebrate, insects and some eubacteria [18, 19]. Family 4 UDGs are a group of prokaryotic thermostable enzymes which excise uracil from DNA strand(s) [20]. Family 5 enzymes are also thermostable and present in archaea and eubacteria, with versatile substrate specificities [21]. Family 6 enzymes (HDG) recognize and excise hypoxanthine in DNA strand(s), exclusively [22].

Firstly isolated from *Xenopus laevis* and human, SMUG1 enzymes show few sequence homologies to other UDG families, but have similar gross structural fold [18].

Later, a genome database search identified SMUG1 orthologs in a few eubacteria lacking family 1 UNG [19, 23]. In comparison to the robust family 1 UNG, SMUG1 enzymes are less active to uracil-containing DNA, but they are versatile with broader substrate specificity, in which they include uracil, xanthine, 5-formyluracil, 5-hydroxyuracil, 5-hydroxymethyluracil and 3, N^4-ethenocytosine as substrates [19, 24-27]. Depending on reaction conditions, SMUG1 presents different substrate preferences. Notably, in the presence of Mg^{2+} under physiological conditions, SMUG1 switches into double-stranded-selective UDG [28]. Through extensive biochemical and genetic analysis, SMUG1 was identified as the principal 5-formyluracil and 5-hydroxylmethyluracil DNA glycoyslase in mammalian systems [24, 26, 27, 29]. Later on, we reported that bacterial SMUG1, as well as human SMUG1, was not only a UDG, but also a xanthine DNA glycosylase [19]. The structures of the bacterial *Geobacter metallireducens* SMUG1 have been resolved [30].

Through a large-scale phylogenetic analysis of UDG superfamily in more than 1,000 completely sequenced genomes, it is shown that family 3 enzymes can be separated into two clades with the traditional family 3 SMUG1 as one of them (Fig. 2.1) [31]. UDG enzymes in another clade are present in eubacteria genus including *Listeria*, *Lactobacillus*, *Streptomyces*, *Amycolatopsis* and *Flavobacteriaceae*. This group of UDGs, which we named as SMUG1-like, is more similar to family 3 SMUG1s and shares a common ancestor with family 1 UNGs. As identified previously [31], a distinct difference between SMUG1-like and traditional SMUG1 is that the “GMNPGP” in motif 1 of SMUG1 is changed to “GSSPAR” in SMUG1-like enzymes. To our knowledge, the

biochemical and enzymatic properties of SMUG1-like enzymes are completely unknown. In this study, we characterized a SMUG1-like DNA glycosylase from *Listeria innocua* (Lin). To our surprise, Lin SMUG1-like glycosylase is a single-stranded UDG with little activity on double-stranded uracil-containing DNA. Mutational analysis indicates similar roles of active site residues corresponding to SMUG1. Interestingly, a double substitution of S67-S68 in Lin SMUG1-like DNA glycosylase by M67-N68 in motif 1 to mimic SMUG1 is able to partially rescue the negative effects of the S67M and S68N single mutants, which implies correlation and co-evolution of the two catalytic residues. The potential physiological role of the SMUG1-like is also discussed.

III. Material and Methods

2.1. Plasmid construction, cloning, expression and purification

The *L. innocua* SMUG1-like gene (GenBank accession number: WP_010991469.1) was amplified by PCR using the forward primer Lin_SMUG1_F (5'- GGG AAT TCC ATA TGG CTA GCA TGA CTG GTG - 3'; the NdeI site is underlined) and the reverse primer Lin_SMUG1_R (5'- CCG CTC GAG CCT CTT TAA AGC ACA -3'; the XhoI site is underlined). The PCR reaction mixture (25 µl) consisted of 8 ng of *L. innocua* genomic DNA, 200 nM forward primer Lin_SMUG1_F and reverse primer Lin_SMUG1_R, 1 x Phusion PCR buffer (New England Biolabs), 200 µM each dNTP, and 1 unit of Phusion DNA polymerase. The PCR procedure included a predenaturation step at 98°C for 5 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 sec, annealing at 50°C for 15 sec and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR product was purified by Gel

DNA Recovery Kit (Zymo Research). Purified PCR product and plasmid pET21a (+) were digested with NdeI and XhoI, purified by Gel DNA Recovery Kit (Zymo research) and ligated according to the manufacturer's instructional manual. The ligation mixture was transformed into *E. coli* strain DH5α competent cells prepared by electroporation. The sequence of the *L. innocua* SMUG1-like gene in the resulting plasmid (pET21a (+)-Lin-SMUG1) was confirmed by DNA sequencing.

The resulting plasmid with wild-type SMUG1-like was used as the template plasmid for all other SMUG1-like mutants. Amplification of mutant DNA and DpnI mediated site-directed mutagenesis procedures were carried out as previously described with modification by using primers carrying the desired mutations [32, 33]. To express the C-terminal His-6-tagged wild-type and mutant Lin SMUG1-like glycosylase gene, the recombinant plasmids were transformed into *E. coli* strain BL21 (DE3 Δ*slyD* Δ*mug* Δ*udg* Δ*nfi* Δ*nth* Δ*ndk*) by electroporation. Protein expression and purification were carried out as previously described [22].

2.2. DNA glycosylase activity assay

The sequence of the oligonucleotides used for DNA glycosylase activity assay is shown in Fig. 2.2B, and prepared as previously described [19]. DNA glycosylase cleavage assays for Lin SMUG1-like proteins were performed at 42°C for 2 h in a 10 µl reaction mixture containing 10 nM oligonucleotide substrate, 20 µM glycosylase protein unless noted otherwise, 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM EDTA and 1 mM DTT. The resulting abasic sites were cleaved by incubation at 95°C for 5 min after adding 1 µl of 1 N NaOH. To quantify cleavage products and remaining substrates, the

reaction mixtures were analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

2.3. Enzyme kinetic analysis

Enzyme kinetic analysis was performed with Lin SMUG1-like glycosylase by incubating different concentrations of respective proteins (1 μM to 100 μM) with 10 nM substrate. Due to the low catalytic efficiency of Lin SMUG1-like, we determined the kinetic parameters under the conditions that enzyme concentration was in excess of substrate concentration as previously described [34]. The linear increase of k_{obs} over enzyme concentrations precluded acquisition of individual k_2 or K_m, but allowed for determination of k_2/K_m [34]. The k_2 / K_m values were obtained by ploting k_{obs} against total enzyme concentration by linear regression with equation $k_{obs} = \frac{k_2[E]}{K_m}$.

2.4. Phylogenetic analysis

A total of 28 glycosylase protein sequences were retrieved from Genbank and aligned using the multiple sequence alignment program ClustalX2 [35] and a structure-based alignment program PROMALS3D [36]. Subsequently, the resulting alignment was curated manually to align the catalytic motifs. The phylogenetic tree was generated using the neighbor-joining method within the MEGA v6.0 software package [37] to infer the evolutionary history between different proteins.

2.5. Spontaneous mutation frequency assay

E. coli ung and Lin SMUG1-like glycosylase genes (WT and mutant) were amplified by PCR with added KpnI and HindIII sites. The amplicons and pBluescript-II

KS(+) were digested with KpnI and HindIII and ligated. Ligation mixtures were transformed into *E. coli* DH5α to produce constructs pBluescript-II KS (+)-UNG (encoding *E. coli* UNG glycosylase), pBluescript-II KS (+)-WT (encoding Lin SMUG1-like glycosylase) and pBluescript-II KS (+)-S67M (encoding Lin SMUG1-like glycosylase S67M mutant). The sequences of the inserts were confirmed by DNA sequencing. *E. coli* strain MG1655 Δ3 (Δung Δmug Δnth) was transformed with the IPTG inducible constructs pBluescript-II KS (+)-UNG, pBluescript-II KS (+)-WT, pBluescript-II KS (+)-S67M (as control) and pBluescript-II KS (+) (as control), respectively. Single colonies were selected and inoculated into 4 ml liquid LB cultures supplemented with 50 μg/ml ampicillin. After overnight incubation at 30°C, IPTG was added to a final concentration of 0.5 mM. The cultures were incubated at 37 °C for an additional 5 h. For cell counting, the cell cultures were then diluted to 1 x 10^8 cells per ml and mixed with 3 ml of 0.7% soft agar and then plated on LB plates with both 50 μg/ml ampicillin and 100 μg/ml rifampicin. The plates were incubated at 37°C for 48 h and colony numbers on amp^R and rif^R plates were counted, respectively. The mutation frequency was calculated as the results of rif^R colony number per 10^9 amp^R colony number.

Mutation rates to rifampicin resistance (rif^R) were also estimated using fluctuation tests as described [38]. Strains used in spontaneous mutation frequency assay as described above were grown in LB broth supplemented with 50 ug/ml ampicillin, 50 ug/ml kanamycin and 1 mM IPTG. Mutation rates from fluctuation tests were calculated

using the Ma-Sandri-Sarkar maximum likelihood method implemented in the FALCOR web tool (www.mitochondria.org/protocols/FALCOR.html) [39, 40].

2.6. Homology modeling

Position-Specific Iterated BLAST (PSI-BLAST) was used to generate the alignment between the amino acid sequence from chain A of the *Xenopus laevis* SMUG1 structure (PDB 1OE5) and SMUG1-like protein sequence from *L. innocua* (GenBank accession WP_010991469.1), which showed 19% identity and 31% similarity. The resulting sequence alignment and the *Xenopus laevis* SMUG1 structure were input to the MODELLER 9.15 software package [41]. A homology model was constructed for Lin SMUG1-like protein with uracil base. Using the CHARMM (ver. c36b1) package, the complex was refined by harmonic restraints with a force constant of 10 kcal/mol/Å^2 to fix coordinates of amino acids within 10 Å of the uracil base, following by Newton-Raphson minimization of 200 steps to remove van der Waals clashes around uracil. After removing Harmonic restraints, a generalized Born implicit solvent was carried out, and the complex was refined for a further 1000 steps. The coordinates of uracil base were refined using the Autodock 4.6.2 package.

2.7. Mutual information (MI) analysis

The mutual information (MI) analysis was performed as previously described with modifications [34]. Briefly, to encompass the six known families of the UDG superfamily in the analysis, we compiled sequences of the families 2, 3, 5 and 6 and added them to the existing set of families 1 and 4 sequences [34]. Subsequently, a multiple sequence alignment was constructed using MEGA6 [37]. The alignment was

used as the input for the mutual information analysis using the MISTIC web server [42]. The reference sequence selected was the SMUG1-like glycosylase from *Listeria innocua* (GenBank accession WP_010991469.1). The parameters were set to those used previously [34].

IV. Results and Discussion

3.1. SMUG1-like enzyme as UDG

Listeria innocua genome contains a traditional family 1 UNG (GenBank accession number EHN61868.1) and a SMUG1-like glycosylase and two AP endonucleases (GenBank accession numbers WP_003769869.1 and KJR54533.1), indicating that it possesses a complete base excision repair pathway. The SMUG1-like group as represented by the enzyme in *L. innocua* shows high degree of sequence homology to SMUG1 and possesses similar motifs except for noted difference in motif 1 (Figs. 2.1, 2.2A). To determine its potential DNA repair function, we assayed the DNA glycosylase activity using the fluorescently labeled oligonucleotide substrates (Fig. 2.2B). The recombinant protein was cloned, expressed in *E. coli* and purified as shown in Fig. 2.2C. Among the five damaged bases tested, the Lin SMUG1-like enzyme only showed DNA glycosylase activity toward uracil-containing DNA substrate under the conditions that the enzyme was in excess over the substrate (Fig. 2.2D). No enzymatic activity was detected with double-stranded and single-stranded hypoxanthine-, xanthine-, 5-hydroxymethyluracil (hmU)-, 5-hydroxyuracil (OHU)-, 8-oxoguanine-, 8-oxoadenine-, 5,6-dihydroxyuracil (DHU)-, 5-hydroxycytosine-, thymine glycol-, N6-methyladenine-, O6-methylguanine-, and ethenoadenine-containing DNA (Fig. 2.2D and data not shown).

Lin SMUG1-like enzyme did not show binding affinity to these substrates either as measured by gel mobility shift analysis (data not shown). Gme SMUG1 is still active with uracil-containing substrates with 1 nM enzyme and 10 nM substrate [19]. However, Lin SMUG1-like glycosylase required excess enzyme to show detectable activity on uracil-containing substrates, suggesting that the overall enzymatic activity was not as robust as Gme SMUG1 [19]. Unexpectedly, Lin SMUG1-like glycosylase was more active with single-stranded uracil-containing DNA and showed only low-level activity with double-stranded uracil-containing DNA (Fig. 2.2D).

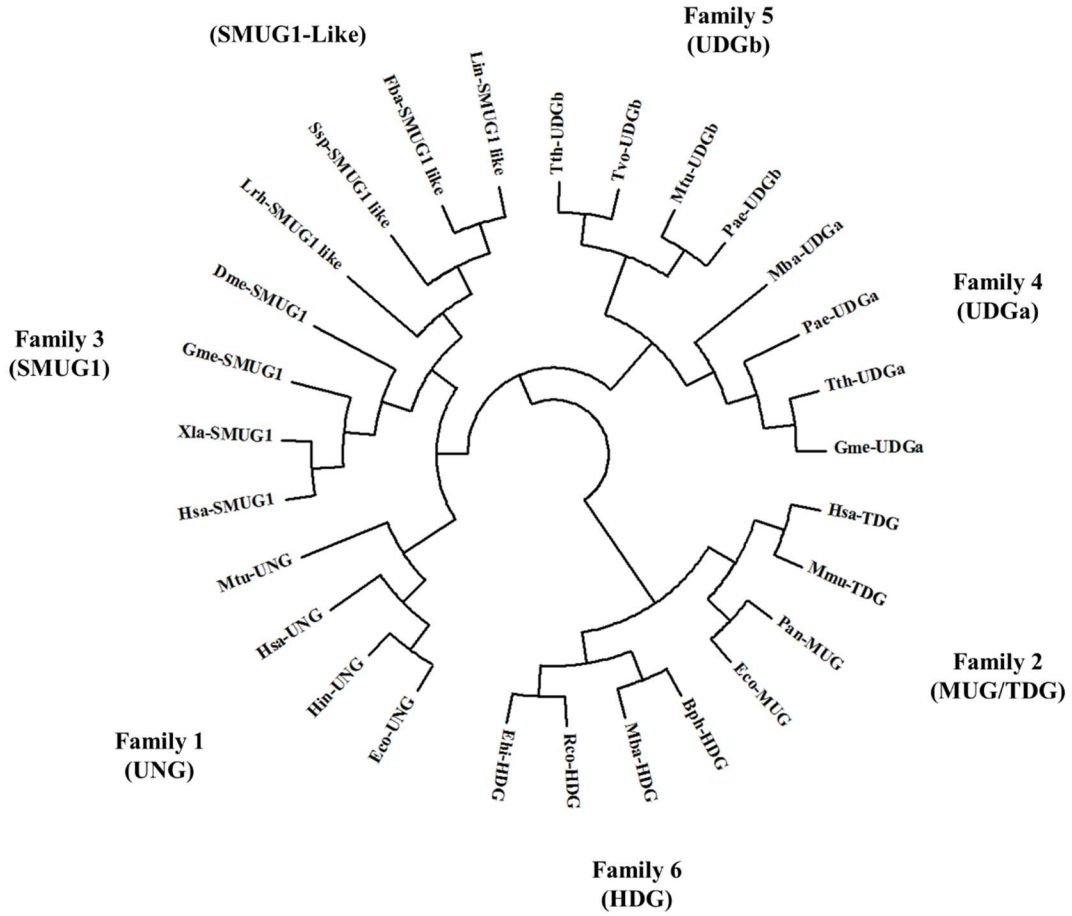


Figure 2. 1 Phylogenetic analysis of UDG superfamily. The phylogenetic analysis was performed using the neighbor-joining method in MEGA 6. GenBank accession numbers are shown after the species names. SMUG1-like glycosylase: Lin*, Listeria innocua* WP_010991469.1; Lrh*, Lactobacillus rhamnosus* WP_014569641.1; Fba*, Firmicutes bacterium CAG:822* WP_021872912.1; Ssp*, Streptomyces sp. NRRL WC-3626* WP_030213602.1. Family 1 (UDG): Eco, *Escherichia coli*, NP_289138; Hin, *Haemophilus influenzae* KR494, YP_008544610.1; Mtu*, Mycobacterium tuberculosis*, WP_003908950.1; Hsa, *Homo sapiens,* NP_003353. Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1; Pan*, Pantoea ananatis LMG 20103*, ADD78558.1; Hsa*, Homo sapiens*, NP_003202; Mmu*, Mus musculus*, XP_003945901.1. Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15,YP_383069; Hsa*, Homo sapiens*, NP_055126; Xla, *Xenopus laevis* AAD17300; Dme*, Drosophila melanogaster*, NP_650609.1. Family 4 (UDGa): Tth, *Thermus thermophilus* HB27, YP_004341.1; Pae*, Pyrobaculum aerophilum str*. IM2, NP_558739.1; Gme*, Geobacter metallireducens* GS-15, YP_006721625.1; Mba*, Methanosarcina barkeri* str. Fusaro, YP_305330.1. Family 5 (UDGb): Tth, *Thermus thermophilus* HB8, YP_144415.1; Pae, *Pyrobaculum aerophilum* str. IM2, NP_559226; Tvo, *Thermoplasma volcanium* GSS1, NP_111346.1; Mtu, *Mycobacterium tuberculosis* H37Rv, P64785 (Rv1259). Family 6 (HDG): Bph*, Burkholderia phymatum* STM815, YP_001858334.1; Mba*, Methanosarcina* barkeri str. Fusaro, YP_304295.1; Rco*, Ricinus communis*, XP_002536323.1; Ehi*, Entamoeba histolytica* HM-1:IMSS, XP_655177.1.

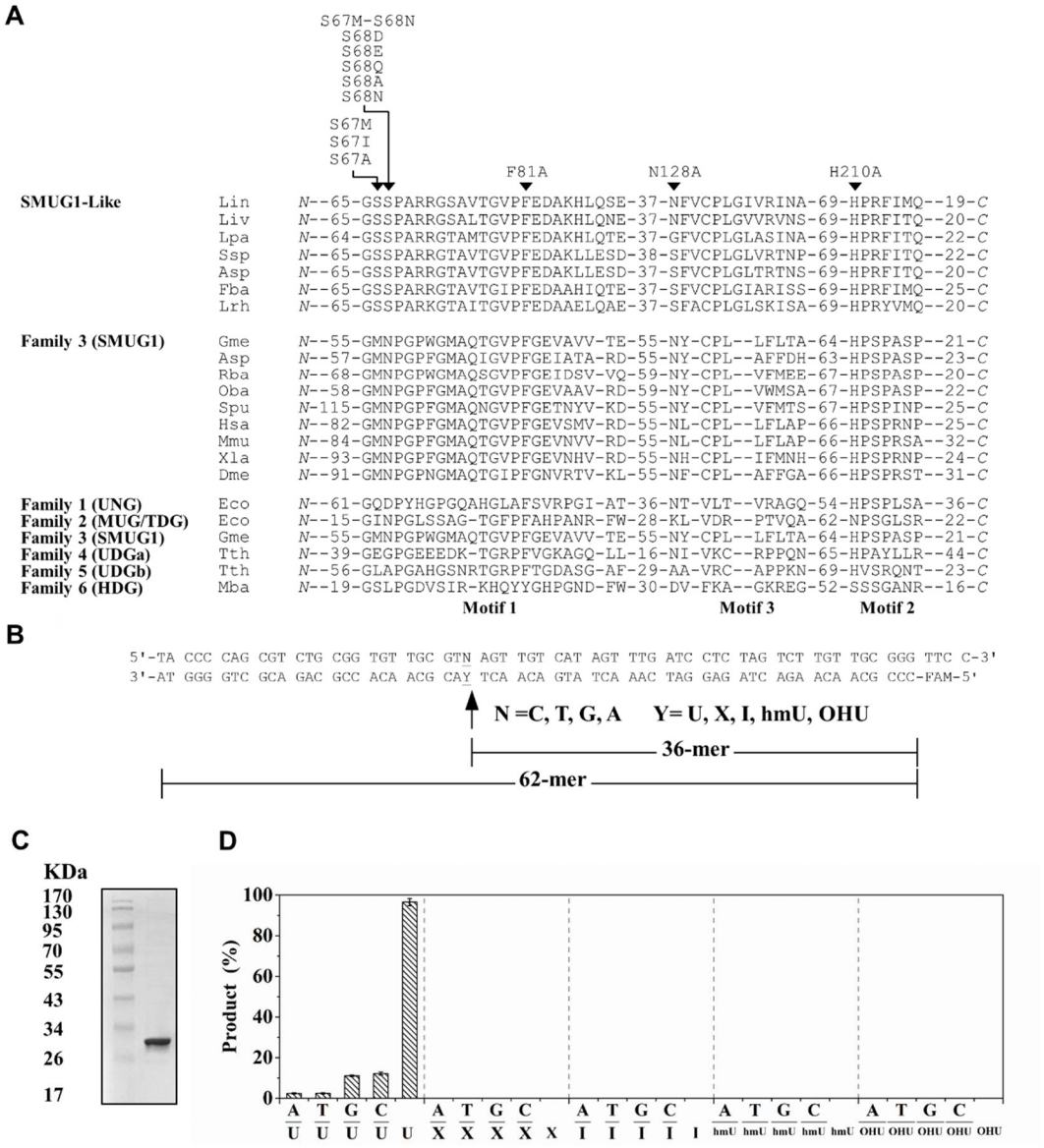


Figure 2. 2 Multiple sequence alignment and DNA glycosylase activity of Lin SMUG1-like glycoyslase. A. Multiple sequence alignment of SMUG1-like and SMUG1 enzymes. GenBank accession numbers are shown after the species names. SMUG-like glycosylase: Lin, *Listeria innocua* WP_010991469.1; Liv, *Listeria ivanovii* WP_025279932.1; Lpa, *Lactobacillus paracasei* WP_016381167.1; Ssp, *Streptomyces sp.* NRRL WC-3626 WP_030213602.1; Asp, Amycolatopsis sp. WP_037336521; Fba*, Firmicutes bacterium CAG:822* WP_021872912.1; Lrh*, Lactobacillus rhamnosus* WP_014569641.1; Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069; Asp, *Azoarcus sp.* BH72, YP_935478; Rba, *Rhodopirellula baltica SH 1*, NP_869403; Oba, *Opitutaceae bacterium* TAV2, ZP_02013615.1; Spu, *Strongylocentrotus purpuratus*, XP_782746.1; Hsa, *Homo sapiens*, NP_055126; Mmu,

Mus musculus, NP_082161; Xla, *Xenopus laevis*, AAD17300; Dme, *Drosophila melanogaster*, NP_650609.1; Family 1 (UNG): Eco, *Escherichia coli*, NP_289138; Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1. Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069. Family 4 (UDGa): Tth, *Thermus thermophilus* HB27, YP_004341.1. Family 5 (UDGb): Tth, *Thermus thermophilus* HB8, YP_144415.1; Family 6 (HDG): Mba, *Methanosarcina barkeri* str. Fusaro, YP_304295.1. B. Sequences of deoxyoligonucleotide substrates. U, uracil; X, xanthine; I, hypoxanthine; hmU, 5-hydroxymethyluracil; OHU, 5-hydroxyuracil. FAM, 6-FAM fluorophore. C. Purified Lin SMUG1-like glycosylase protein as shown in 12% SDS-PAGE gel. D. DNA glycosylase activity of wild type Lin SMUG1-like glycosylase on U-, X-, I-, hmU- and OHU-containing substrates. Cleavage reactions were performed as described in Material and Methods with 20 μM wild type Lin SMUG1-like protein and 10 nM substrate and incubated for 2 hours. Data are the averages of three independent experiments.

To examine whether the strong preference for single-stranded U-containing DNA was simply the outcome of assay conditions, we examined UDG activity under different conditions. The enzyme showed optimal ss UDG activity at pH 7.5; however, the ds UDG activity remained very low across all pH values tested (Fig. 2.3A). Likewise, salt titration with KCl did not change its ss U preference (Fig. 2.3B). Lin SMUG1-like glycosylase had highest activity at 42°C with ss U but still with minimal activity on ds U at all temperatures examined (Fig. 2.3C). The glycosylase activity on ss U was enhanced with Mg^{2+} but suppressed by transition metal ions Co^{2+}, Fe^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+} and Cd^{2+} (Fig. 2.3D). These results indicate that the strong preference for ss U is an intrinsic property of the enzyme rather than a reflection of assay conditions. We tested the affinity of this glycosylase to single-stranded uracil-containing DNA, but it did not show retarded band in gel mobility shift analysis (data not shown). Because the overall enzymatic activity of Lin SMUG1-like glycosylase was low, we measured k_2/K_m under the condition that the enzyme concentration was in excess. This assay method was used previously to

study family 4 UDGa [34]. The k_2/K_m value for ss U was 3.5 x 10^{-6} min^{-1} nM^{-1}, confirming that the overall kinetic efficiency of the enzyme was low (Table 1).

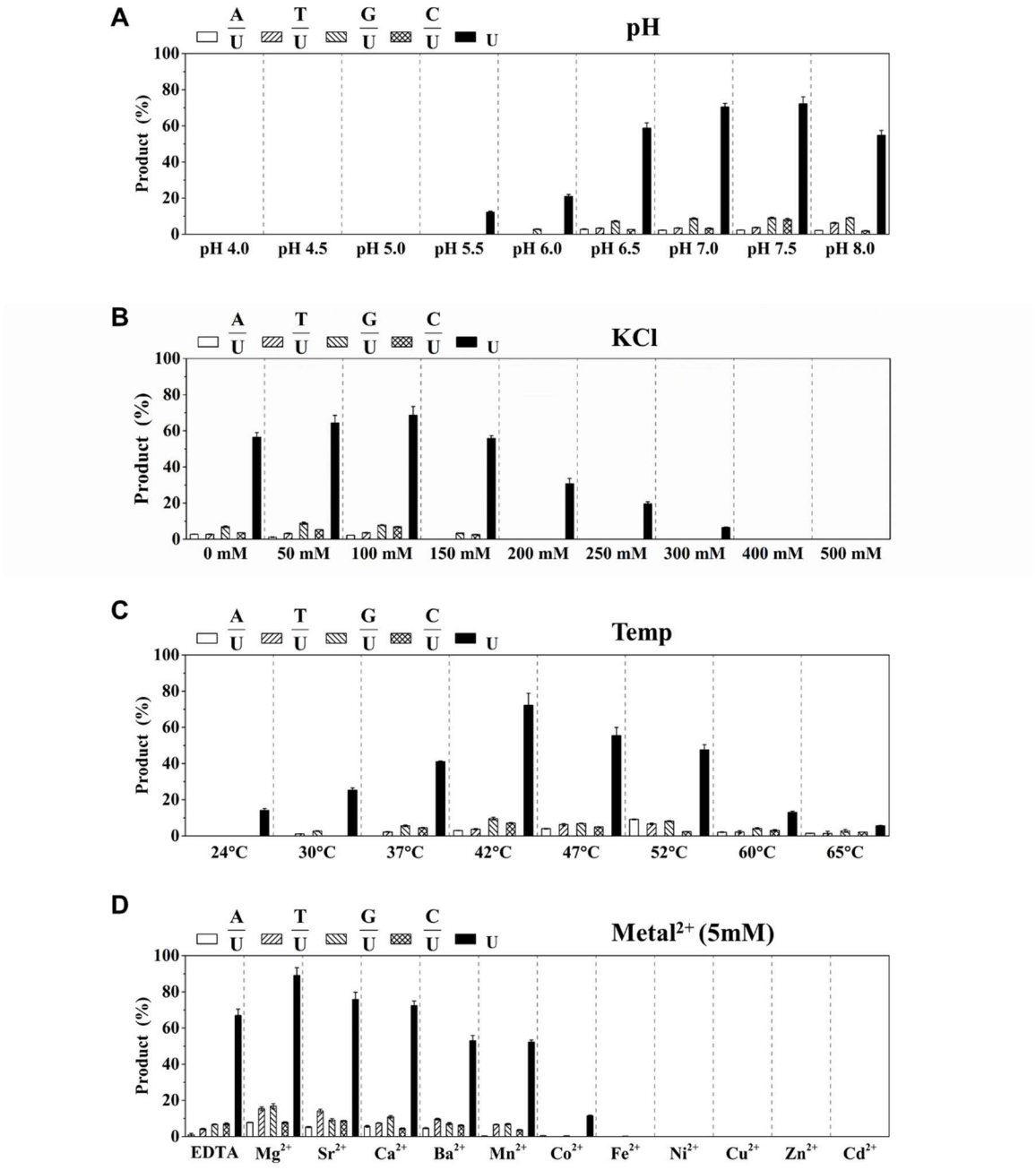


Figure 2. 3 Effects of assay conditions on UDG activity of Lin SMUG1-like glycosylase. DNA glycosylase assays were performed as described in Material and Methods with 5 μM wild type Lin SMUG1-like glycosylase and 10 nM DNA substrate

and incubated for 1 hour. Data are the averages of three independent experiments. A. pH effect. The Tris-HCl buffer was replaced with 20 mM citrate-phosphate universal buffer at the indicated pH value. 4.0-8.0. B. salt effect. KCl was added to the reaction mixtures in the range of 0 to 500 mM. C. Temperature effect. The reaction mixtures were incubated at indicated temperatures before quenching. D. Divalent metal ion effect. Divalent metal ions in the final concentration of 5 mM were supplemented to the reaction mixtures.

Table 2. 1 Kinetic constants of Lin SMUG1-like WT and mutant proteins for excision of uracil[a]

Lin SMUG1	Substrate	k_2/K_m (min^{-1} nM^{-1})
WT	ss U	3.5×10^{-6}
	19 bubble	1.5×10^{-6}
S67M	ss U	N.A.[b]
	19 bubble	N.A.
S68N	ss U	6.4×10^{-8}
	19 bubble	N.A.
S67M-S68N	ss U	1.4×10^{-7}
	19 bubble	N.A.

[a] Kinetic analysis was carried out as described in Material and Methods. Data are the averages of three independent experiments.
[b] N.A. No activity detected under the assay conditions.

To further characterize the strong preference to ss U, we tested the UDG activity toward different bubble structures. A bubble structure is a double-stranded DNA containing an unpaired segment (Fig. 2.4). We designed 7-nucleotide, 13-nucleotide and 19-nucleotide bubbles (Fig. 2.4A-B). A variant of the 19-nt bubble was also included by placing the uracil base 3-nt from the nearest double strand junction (Fig. 2.4B, 19-nt shift Bubbled DNA). The UDG activity followed the order of ss U > 19-nt bubble > 13-nt

bubble > 7-nt bubble > 19-nt shift bubble (Fig. 2.4C). The k_2/K_m value of the 19-nt bubble is 2-fold lower than that of the ss U substrate (Table 2.1). These results revealed that the closer to the double strand junction the uracil is, the lower the UDG activity. Hence, Lin SMUG-like glycosylase is adapted to act on ss U-containing DNA.

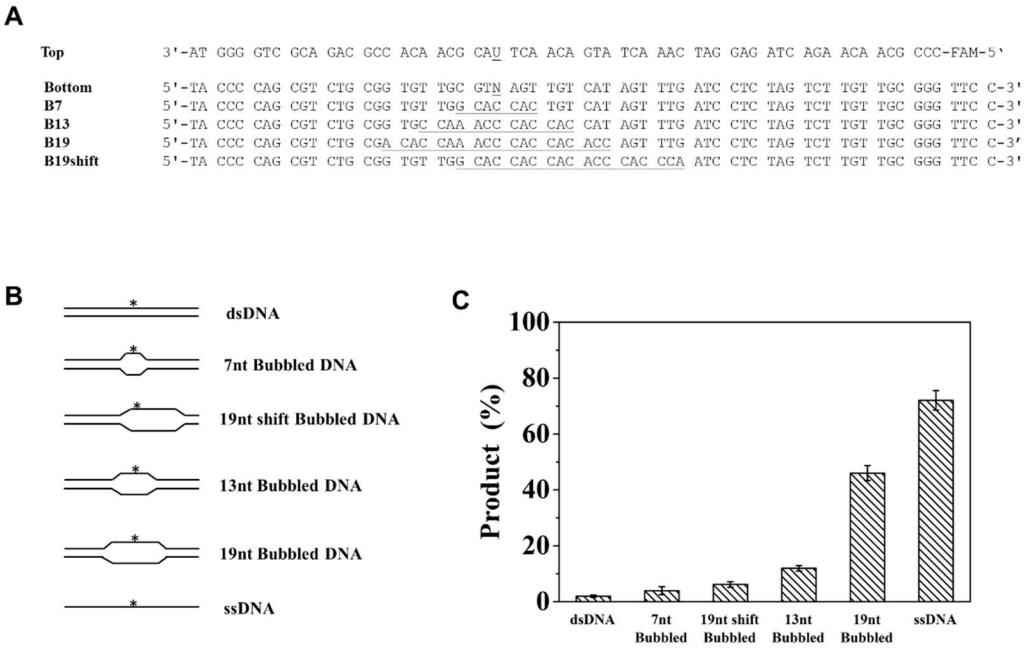


Figure 2. 4 UDG activities of Lin SMUG1-like glycosylase on bubble-containing DNA substrates. A. Sequence of bubble substrates. B. Schematic illustration of bubble substrates. C. UDG activity of Lin SMUG1-like glycosylase on bubble-containing DNA substrates. Cleavage reactions were performed as described in Material and Methods with 5 μM Lin SMUG1-like and 10 nM substrate and incubated for 1 hour. Data are the average of three independent experiments.

3.2. In vivo analysis

To test whether the UDG activity could play a role *in vivo*, we initially measured mutation frequencies using the rifampicin-based assay [43, 44]. *E. coli* family 1 UNG is well known for its extremely robust UDG activity. As expected, the presence of *E. coli* UNG reduced the mutation frequency to 32 Rif^R colonies per 10^9 viable cells (Table 2.2).

In the presence of Lin SMUG1-like glycosylase, the mutation frequency reached 464 Rif^R colonies per 10^9 viable cells (Table 2.2), which was about 15-fold higher than in the presence of *E. coli* UNG but 2-fold lower than its absence (Table 2.2, compare pBS-WT with pBS-S67M and pBS). To more precisely measure the mutation rates per cell division, we performed fluctuation tests as previously described [38]. The presence of *E. coli* UNG reduced the mutation frequency to 24 Rif^R colonies per 10^9 viable cells (Table 2.2). In the presence of Lin SMUG1-like glycosylase, the mutation frequency reached 146 Rif^R colonies per 10^9 viable cells (Table 2.2), which was about 6-fold higher than in the presence of *E. coli* UNG but about 2-fold lower than its absence (Table 2.2, compare pBS-WT with pBS-S67M and pBS). These results suggest that Lin SMUG1-like glycosylase can reduce mutation *in vivo* although not as efficiently as family 1 *E. coli* UNG. The single-stranded UDG activity may help repair of base damage occurring during DNA replication and transcription, as previously suggested [45].

Table 2. 2 Antimutator effect of the Lin SMUG1 like glycosylase[a]

Plasmid	pBS-UNG	pBS-WT	pBS-S67M	pBS
Mutation Frequency ($Rif^R/10^9$ viable cells)[b]	32	464	939	950
Mutation Frequency ($Rif^R/10^9$ viable cells per generation)[c]	24	146	247	254

[a]: pBS-UNG, pBluescript containing *E. coli ung* gene; pBS-WT, pBluescript containing Lin SMUG1-like glycosylase gene; pBS-S67M, pBluescript containing the S67M mutant of the Lin SMUG1-like glycosylase gene; pBS, pBluescript alone.
[b]: The average mutation frequency was determined from at least 3 independent experiments.
[c]: Each mutation frequency was determined from 48 to 60 independent cultures.

3.3. Site-directed mutagenesis of Lin SMUG1-like glycosylase in motifs 1, 2 and 3

UDG enzymes have three conserved motifs that are important for their DNA glycosylase activity (Fig. 2.2A). As shown in the multiple sequence alignment, SMUG1-like enzymes share some sequence conservation in the three motifs, in particular F81 in motif 1, N128A in motif 3 and H210 in motif 2 (Fig. 2.2A). In addition, as described above, a unique feature of the SMUG1-like enzymes is the substitution of GMNP with GSSP in motif 1. To investigate the role of these residues in Lin SMUG1-like enzyme, we substituted F81, N128 and H210 with alanine, and S67 and S68 with several residues (Fig. 2.2A). Alanine substitution at F81, N128 and H210A positions rendered the enzyme much less active compared with the WT enzyme (Table 2.3). Similar effects of these substitutions in SMUG1 enzymes were reported previously [23]. To understand the mutational effects, we modeled the structure of Lin SMUG1-like glycosylase using Xenopus SMUG1 as a template. Molecular modeling indicates that these three residues

interact with uracil in a way similar to what has been observed in family 3 SMUG1 [46, 47]. The highly conserved F81 in motif 1 stacks with uracil to stabilize the flipped out base (Fig. 2.5A). Likewise, F109 in Xenopus SMUG1 and F98 in human SMUG1 play a similar role (Fig. 2.5B) [46, 47]. Situated in the same location as N174 in Xenopus SMUG1 and N163 in Human SMUG1 [46, 47], N128 in Lin SMUG1-like glycosylase could interact with N3 and O4 of the uracil base (Fig. 2.5A-B). The first histidine residue in motif 2 is proposed to form a hydrogen bond with O2 of uracil to stabilize the negatively charged uracil leaving group in family 1 UNG, family 4 UDGa and family 5 UDGb [21, 34, 48]. It is likely that H210 plays a similar catalytic role in Lin SMUG1-like glycosylase (Fig. 2.5A-B).

In Lin SMUG1-like glycosylase, S67 is located in the similar position as M95 in Xenopus SMUG1 (Figs. 2.2A and 2.5C-D). The mainchain amino group of M95 or S67 interacts with the O2 of uracil (Fig. 2.5C-D). Substitution of S67 with Ala, Ile and Met all resulted in loss of glycosylase activity (Table 2.3). Even though the equivalent position in family 2 *E. coli* MUG and family 3 Xenopus SMUG1 is Ile and Met, respectively, the S67I and S67M mutants were inactive. These results suggest that although Ile and Met residues are compatible in family 2 and family 3 enzymes, they are not in SMUG1-like enzymes.

In Lin SMUG1-like glycosylase, S68 is located in the similar position as N96 in Xenopus SMUG1 (Figs. 2.2A and 2.5C-D). In *E. coli* MUG, the mainchain amino group and the sidechain amide group of N18 located in the equivalent position are proposed to position a water molecule for catalysis [11, 49]. Likewise, N96 in Xenopus SMUG1 and

S68 in Lin SMUG1-like glycosylase could play a similar catalytic role by using the mainchain and sidechain to position the catalytic water (Fig. 2.5C-D). Substitution of S68 with Ala resulted in loss of glycosylase activity (Table 2.3). However, substitution of S68 with Asn, Asp, Glu and Gln retained UDG activity to various degrees (Table 2.3). These results may indicate the need of functional sidechain to position the water molecule (Fig. 2.5C). Overall, Lin SMUG1-like glycosylase adopts a similar catalytic mechanism as SMUG1, in which H210 is responsible for promoting the departure of the uracil leaving group and S68 is needed to position a water molecule to attack the N-glycosidic bond (Fig. 2.5E).

To test whether S67 and S68 are correlated, we constructed an S67M-S68N double mutation. S67M-S68N mutant turned out to be more active than the S67M or S68N single mutant with single-stranded uracil-containing DNA (Table 2.3), suggesting that these two positions are correlated. No detectable glycosylase activity was found with double-stranded uracil-containing DNA. To more quantitatively compare the mutational effects of the single mutants, double mutant with the WT Lin SMUG1-like glycosylase, we determined the enzyme kinetics parameters. S68N single mutation reduced the k_2/K_m value by two orders of magnitude (Table 2.1). Even though S67M-S68N double mutation reduced the k_2/K_m value by 25-fold as compared with the wild type enzyme, it increased the k_2/K_m value by 2-fold over the S68N mutation (Table 2.1). Compared with S67M single mutation, even though S67M did not show any detectable UDG activity under the assay conditions, adding S68N mutation to it apparently reactivated the UDG activity (Table 2.1). These results indicate that S67M and S68N can work as a correlated pair in

Lin SMUG1-like glycosylase, although less efficiently than the natural S67-S68 pair. To assess whether the enhanced glycosylase activity on ss U was due to a relaxation of specificity, we examined the activity using double-stranded and single-stranded hypoxanthine-, xanthine-, 5-hydroxymethyluracil (hmU)-, 5-hydroxyuracil (OHU)-, 8-oxoguanine-, 8-oxoadenine-, 5,6-dihydroxyuracil (DHU)-, 5-hydroxycytosine-, thymine glycol-, N6-methyladenine-, O6-methylguanine-, and ethenoadenine-containing DNA with 20 μM enzyme and 10 nM substrate. No detectable activity was found under the assay conditions, suggesting that the S67M-S68N doublet mutant retained its specificity.

Table 2. 3 Gycosylase activity of Lin SMUG1-like glycosylase on ss U-containing DNA[a]

	A U	T U	T U	T U	U
wild type	2	3	7	8	100
S67A	0	0	0	0	0
S67I	0	0	0	0	0
S67M	0	0	0	0	0
S68A	0	0	0	0	0
S68N	0	0	0	0	11
S68D	0	0	0	0	17
S68E	0	0	0	0	38
S68Q	0	0	0	0	7
S67M-S68N	0	0	0	0	48
F81A	0	0	0	0	21
N128A	0	0	0	0	14
H210A	0	0	0	0	21

[a]: The reactions were performed as described in Material and Methods with 20 μM Lin SMUG1-like enzyme and 10 nM DNA substrate and incubated for 2 hours. Data are the averages of three independent experiments.

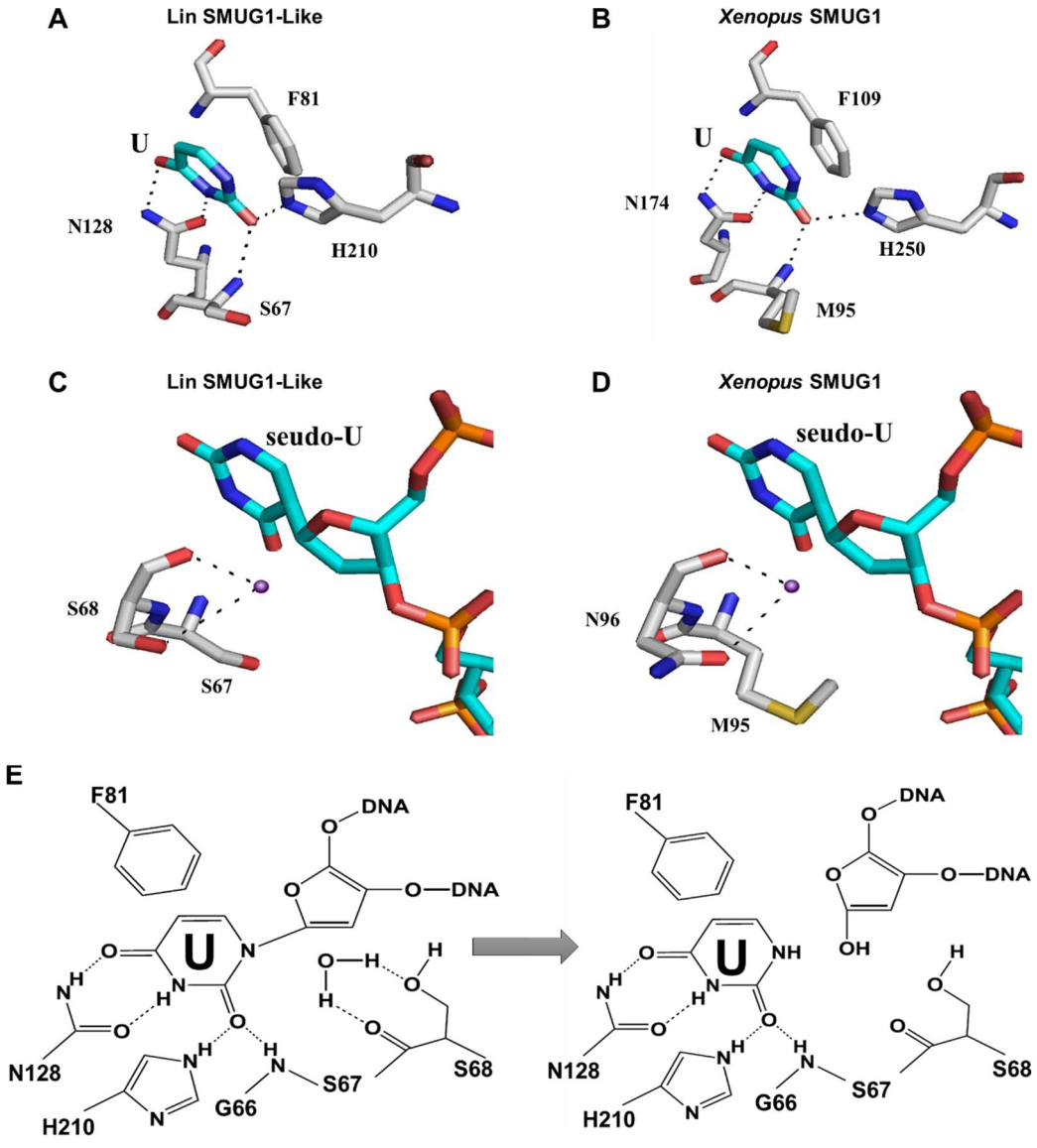


Figure 2. 5 Interactions between Lin SMUG1-like glycosylase and uracil-containing DNA. Uracil is shown in cyan. Hydrogen bonds are shown as dash lines. A. Interactions between Lin SMUG1-like glycosylase and uracil base. Lin SMUG1-like glycosylase structure was modeled based on Xenopus SMUG1 structure (PDB code 1OE5). B. Interactions between Xenopus SMUG1 glycosylase (PDB code 1OE5) and uracil base. C. Modeled interactions between S67 and S68 of Lin SMUG1-like glycosylase and uracil base. A water molecule is shown as a purple circle. The DNA with seudo-U base is taken from PDB 1EMH [50]. D. Modeled interactions between M95 and N96 of Xenopus SMUG1 glycosylase and uracil base. A water molecule is shown as a purple circle. The DNA with seudo-U base is taken from PDB 1EMH. E. Hypothetical catalytic mechanism of Lin SMUG1-mediated hydrolysis of the N-glycosidic bond.

3.4. Mutual information analysis

Previously, we used mutual information-based method to investigate amino acid correlation in family 4 UDGa and family 1 UNG enzymes [34]. Applying similar analysis to the SMUG1-like enzymes using six families in the UDG superfamily, we found that S67 and S68 in Lin SMUG1-like DNA glycosylase were also correlated (Fig. 2.6). This is consistent with our experimental results showing S67M-S68N is more active as a UDG than either S67M or S68N single mutant (Table 2.1). The equivalent positions to S67-S68 in family 4 Tth UDGa are the E41-G42 pair. Similar to the results reported here, the E41Q-G42D double mutant is more active than E41Q or G42D single mutant [34]. Taken together, these results underscore the correlated nature of the two neighboring residues in different UDG families. As explained above, these two positions play important roles in protein-DNA interactions and catalysis. Understandably, this amino acid doublet may work in concert to carry out its catalytic function in UDG enzymes.

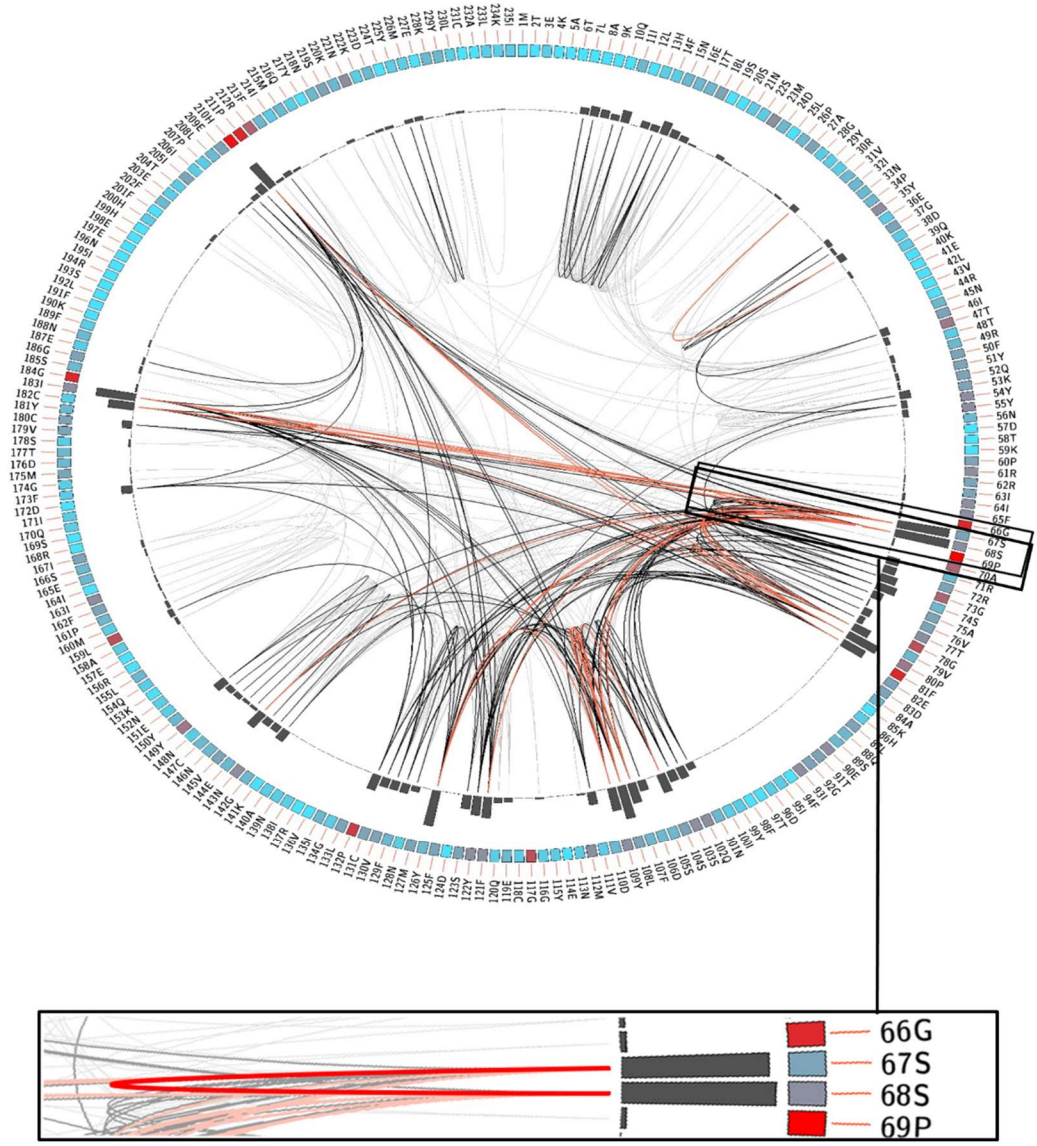


Figure 2. 6 Circos diagram of Lin SMUG1-like glycosylase. The circular representation of the multiple sequence alignment using Lin SMUG1-like glycosylase as the reference sequence. The diagram contains the amino acid residue positions and residue identities mapped to the reference sequence. The square box below each residue represents the level of conservation ranging from red (highly conserved) to blue (less conserved). The bars in the histogram represent the co-evolutionary correlations from the mutual information analysis with a value higher than 6.5 [42]. The connecting lines between residue pairs follow a color scheme for ranking correlation between positions in the multiple sequence alignment where red indicates the top 5%, black between 95% and 70% and gray the remaining interactions.

3.5. Concluding remarks

Even though *E. coli* UNG, as the first discovered DNA glycosylase, is a narrow specificity enzyme acting on uracil-containing DNA [6], now we know that enzymes in UDG superfamily are quite diverse in their specificities. Enzymes in families 2, 3, 4, and 6 can work on purine deamination damage such as xanthine or hypoxanthine DNA glycosylase. As for the UDG activity, family 1 and 4 enzymes can all remove uracil from double- and single-stranded DNA [34]. Some family 2 and 5 enzymes prefer double-stranded DNA [13, 21]. Family 3 SMUG1 enzymes were first identified as a single-stranded selective UDG [18], yet later UDG activity on double-stranded DNA was detected [19, 23, 25, 47, 51]. Unlike family 3 SMUG1 enzymes, the *Listeria* SMUG1-like UDG indeed prefers single-stranded uracil-containing DNA under all assay conditions we tested. The genome of *L. innocua* is 3.09 million bp long with an average G+C content of 37.4% [52]. The weak stacking interactions between A/T pairs can cause higher frequency of opening up DNA to create intermittent single-stranded domains, so-called DNA bubbles and once opened longer lifetime [53, 54]. Spontaneous cytosine deamination occuring in single-stranded DNA is 100-fold greater in comparison to that in the DNA duplex [55, 56]. *Listeria* can also live in high temperature environment [57], at which base deamination is accelerated. The SMUG1-like DNA glycosylase with strong preference for single-stranded DNA may be needed to counter the mutagenic cytosine deamination and maintain the integrity of this low G/C genome. The SMUG1-like enzyme from *Listeria* represents the first example of a single-stranded DNA glycosylase in UDG superfamily.

V. **References**

1. Brynolf, K., Eliasson, R. & Reichard, P. (1978) Formation of Okazaki fragments in polyoma DNA synthesis caused by misincorporation of uracil, *Cell.* **13**, 573-80.

2. Krokan, H. E., Drablos, F. & Slupphaug, G. (2002) Uracil in DNA--occurrence, consequences and repair, *Oncogene.* **21**, 8935-48.

3. Lindahl, T. (1993) Instability and decay of the primary structure of DNA, *Nature.* **362**, 709-15.

4. Neuberger, M. S., Harris, R. S., Di Noia, J. & Petersen-Mahrt, S. K. (2003) Immunity through DNA deamination, *Trends Biochem Sci.* **28**, 305-12.

5. Delker, R. K. & Papavasiliou, F. N. (2013) You break it, you fix it: functions for AID downstream of deamination, *Nat Immunol.* **14**, 1112-4.

6. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues, *Proceedings of the National Academy of Sciences of the United States of America.* **71**, 3649-53.

7. Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily, *Mutation research.* **460**, 165-81.

8. Parikh, S. S., Putnam, C. D. & Tainer, J. A. (2000) Lessons learned from structural results on uracil-DNA glycosylase, *Mutat Res.* **460**, 183-99.

9. Gallinari, P. & Jiricny, J. (1996) A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase, *Nature.* **383**, 735-8.

10. Lee, H. W., Brice, A. R., Wright, C. B., Dominy, B. N. & Cao, W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase, *The Journal of biological chemistry.* **285**, 41483-90.

11. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions, *Cell.* **92**, 117-29.

12. Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L. & Pearl, L. H. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex, *EMBO J.* **18**, 6599-609.

13. Lee, D. H., Liu, Y., Lee, H. W., Xia, B., Brice, A. R., Park, S. H., Balduf, H., Dominy, B. N. & Cao, W. (2015) A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs, *Nucleic acids research.* **43**, 1081-9.

14. Hang, B., Downing, G., Guliaev, A. B. & Singer, B. (2002) Novel activity of Escherichia coli mismatch uracil-DNA glycosylase (Mug) excising 8-(hydroxymethyl)-3,N4-ethenocytosine, a potential product resulting from glycidaldehyde reaction, *Biochemistry.* **41**, 2158-65.

15. Saparbaev, M. & Laval, J. (1998) 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 8508-13.

16. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C. & Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, *Science.* **333**, 1303-7.

17. Zhang, P., Huang, B., Xu, X. & Sessa, W. C. (2013) Ten-eleven translocation (Tet) and thymine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a, *Biochem Biophys Res Commun.* **437**, 368-73.

18. Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W. & Verdine, G. L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors, *Current biology : CB.* **9**, 174-85.

19. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N. & Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG, *Journal of molecular biology.* **385**, 761-78.

20. Haas, B. J., Sandigursky, M., Tainer, J. A., Franklin, W. A. & Cunningham, R. P. (1999) Purification and characterization of Thermotoga maritima endonuclease IV, a thermostable apurinic/apyrimidinic endonuclease and 3'-repair diesterase, *Journal of bacteriology.* **181**, 2834-9.

21. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N. & Cao, W. (2014) Specificity and catalytic mechanism in family 5 uracil DNA glycosylase, *The Journal of biological chemistry.* **289**, 18413-26.

22. Lee, H. W., Dominy, B. N. & Cao, W. (2011) New Family of Deamination Repair Enzymes in Uracil-DNA Glycosylase Superfamily, *J Biol Chem.* **286**, 31282-7.

23. Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E. & Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms, *Nucleic acids research.* **35**, 3879-92.

24. Boorstein, R. J., Cummings, A., Jr., Marenstein, D. R., Chan, M. K., Ma, Y., Neubert, T. A., Brown, S. M. & Teebor, G. W. (2001) Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1, *The Journal of biological chemistry.* **276**, 41991-7.

25. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E. & Slupphaug, G. (2002) hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup, *J Biol Chem.* **277**, 39926-36.

26. Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A. & Ide, H. (2003) Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions, *Biochemistry.* **42**, 5003-12.

27. Matsubara, M., Masaoka, A., Tanaka, T., Miyano, T., Kato, N., Terato, H., Ohyama, Y., Iwai, S. & Ide, H. (2003) Mammalian 5-formyluracil-DNA glycosylase. 1. Identification and characterization of a novel activity that releases 5-formyluracil from DNA, *Biochemistry.* **42**, 4993-5002.

28. Doseth, B., Ekre, C., Slupphaug, G., Krokan, H. E. & Kavli, B. (2012) Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil, *DNA repair.* **11**, 587-93.

29. Kemmerich, K., Dingler, F. A., Rada, C. & Neuberger, M. S. (2012) Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice, *Nucleic acids research.* **40**, 6016-25.

30. Zhang, Z., Shen, J., Yang, Y., Li, J., Cao, W. & Xie, W. (2016) Structural Basis of Substrate Specificity in Geobacter metallireducens SMUG1, *ACS chemical biology.* **11**, 1729-36.

31. Lucas-Lledo, J. I., Maddamsetti, R. & Lynch, M. (2011) Phylogenomic analysis of the uracil-DNA glycosylase superfamily, *Molecular biology and evolution.* **28**, 1307-17.

32. Fisher, C. L. & Pei, G. K. (1997) Modification of a PCR-based site-directed mutagenesis method, *BioTechniques.* **23**, 570-1, 574.

33. Pang, P., Yang, Y., Li, J., Wang, Z., Cao, W. & Xie, W. (2017) SMUG2 DNA glycosylase from Pedobacter heparinus as a new subfamily of the UDG superfamily, *The Biochemical journal.* **474**, 923-938.

34. Xia, B., Liu, Y., Guevara, J., Li, J., Jilich, C., Yang, Y., Wang, L., Dominy, B. N. & Cao, W. (2017) Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily, *Sci Rep.* **7**, 45978.

35. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics.* **23**, 2947-8.

36. Pei, J., Kim, B. H. & Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments, *Nucleic acids research.* **36**, 2295-300.

37. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0, *Molecular biology and evolution.* **30**, 2725-9.

38. Foster, P. L. (2006) Methods for determining spontaneous mutation rates, *Methods in enzymology.* **409**, 195-213.

39. Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis, *Bioinformatics.* **25**, 1564-5.

40. Sarkar, S., Ma, W. T. & Sandri, G. H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants, *Genetica.* **85**, 173-9.

41. Webb, B. & Sali, A. (2014) Comparative Protein Structure Modeling Using MODELLER, *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]*. **47**, 5 6 1-32.

42. Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M. & Marino Buslje, C. (2013) MISTIC: Mutual information server to infer coevolution, *Nucleic acids research.* **41**, W8-14.

43. Otterlei, M., Kavli, B., Standal, R., Skjelbred, C., Bharati, S. & Krokan, H. E. (2000) Repair of chromosomal abasic sites in vivo involves at least three different repair pathways, *EMBO J.* **19**, 5542-51.

44. Mi, R., Alford-Zappala, M., Kow, Y. W., Cunningham, R. P. & Cao, W. (2012) Human endonuclease V as a repair enzyme for DNA deamination, *Mutat Res.* **735**, 12-18.

45. Dou, H., Mitra, S. & Hazra, T. K. (2003) Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2, *J Biol Chem.* **278**, 49679-84.

46. Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L. & Pearl, L. H. (2003) Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1, *Mol Cell.* **11**, 1647-59.

47. Matsubara, M., Tanaka, T., Terato, H., Ohmae, E., Izumi, S., Katayanagi, K. & Ide, H. (2004) Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase, *Nucleic acids research.* **32**, 5291-302.

48. Drohat, A. C. & Stivers, J. T. (2000) Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2, *Biochemistry.* **39**, 11865-75.

49. Barrett, T. E., Savva, R., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. (1998) Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct, *Nature structural biology.* **5**, 697-701.

50. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M. & Tainer, J. A. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects, *Proceedings of the National Academy of Sciences of the United States of America.* **97**, 5083-8.

51. Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L. & Lindahl, T. (2001) Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase, *EMBO J.* **20**, 4278-86.

52. Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K. D., Fsihi, H., Garcia-del Portillo, F., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Perez-Diaz, J. C., Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. & Cossart, P. (2001) Comparative genomics of Listeria species, *Science.* **294**, 849-52.

53. Ambjornsson, T., Banik, S. K., Krichevsky, O. & Metzler, R. (2007) Breathing dynamics in heteropolymer DNA, *Biophys J.* **92**, 2674-84.

54. Ambjornsson, T., Banik, S. K., Krichevsky, O. & Metzler, R. (2006) Sequence sensitivity of breathing dynamics in heteropolymer DNA, *Phys Rev Lett.* **97**, 128105.

55. Frederico, L. A., Kunkel, T. A. & Shaw, B. R. (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy, *Biochemistry.* **29**, 2532-7.

56. Lindahl, T. & Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid, *Biochemistry.* **13**, 3405-10.

57. Fairchild, T. M. & Foegeding, P. M. (1993) A proposed nonpathogenic biological indicator for thermal inactivation of Listeria monocytogenes, *Applied and environmental microbiology.* **59**, 1247-50.

Note:

This chapter has been published as a peer reviewed paper:
Jing Li*, Ye Yang*, Jose Guevara, Liangjiang Wang and Weiguo Cao. Identification of a prototypical single-stranded uracil DNA glycosylase from *Listeria innocua*. **DNA Repair (Amst).** 2017 57:107-115. * These authors contribute equally

CHAPTER THREE

AN UNCONVENTIONAL FAMILY 1 URACIL DNA GLYCOSYLASE IN *NITRATIFRACTOR SALSUGINIS*

I. Abstract

The uracil DNA glycosylase superfamily consists of at least six families with a diverse specificity towards DNA base damage. Family 1 UNG exhibits exclusive specificity on uracil-containing DNA. Here, we report a family 1 UNG homolog from *Nitratifractor salsuginis* with distinct biochemical features that differentiate it from conventional family 1 UNGs. Globally, the crystal structure of *N. salsuginis* UNG shows a few additional secondary structural elements. Biochemical and enzyme kinetic analysis, coupled with structural determination, molecular modeling and molecular dynamics simulations, shows that *N. salsuginis* UNG contains a salt bridge network that plays an important role in DNA backbone interactions. Disruption of the amino acid residues involved in the salt bridge greatly impedes the enzymatic activity. A tyrosine residue in motif 1 (GQDPY) is one of the distinct sequence features setting family 1 UNG apart from other families. The crystal structure of Y81G mutant indicates that several subtle changes may account for its inactivity. Unlike the conventional family 1 UNG enzymes, *N. salsuginis* UNG is not inhibited by Ugi, a potent inhibitor specific for family 1 UNG. This study underscores the diversity of paths that a uracil DNA glycosylase may take to acquire its unique structural and biochemical properties during evolution.

II. Introduction

DNA is subjected to a variety of damage caused by environmental assault and endogenous agents. DNA bases are subject to deamination by hydrolytic or oxidative reactions due to the reactivity of the exocyclic amino groups [1-3]. Cytosine (C), guanine (G), adenine (A), and 5-methylcytosine (mC) can be deaminated to generate uracil (U), hypoxanthine (I), xanthine (X), and thymine (T), respectively. Because of the amino to keto conversion, base deamination alters the hydrogen bonding properties of the damaged bases from a hydrogen donor to a hydrogen acceptor. Consequently, mutations can be generated during DNA replication. Uracil may also appear in DNA as an A/U base pair due to misincorporation of dUTP into the genome by DNA polymerase.

Uracil DNA glycosylase (UDG) initiates the base excision repair (BER) for the repair of uracil or other types of base damage. At least six families are identified in UDG superfamily. Family 1 UNGs, which are common in bacteria and eukaryotes, show rather narrow specificity towards uracil and its derivatives [4, 5]. Family 2 MUG/TDG enzymes, also found in bacteria and eukaryotes, have much broader specificity with some members showing activity toward all deaminated bases [6-9]. Human TDG is now known as part of the demethylation system due to its DNA glycosylase activity for formyl-C (fC) and carboxyl-C (caC) [10, 11], indicating a functional transition beyond DNA repair. Family 3 SMUG1 enzymes, found in vertebrates and bacteria, are known as uracil DNA glycosylases and xanthine DNA glycosylases (XDG) [12, 13]. Family 4 UDGa enzymes, found in bacteria and archaea, have a narrow substrate specificity toward uracil [14]. Family 5 UDGb glycosylases, also found in bacteria and archaea, can act on uracil,

hypoxanthine (deamination product of adenine), xanthine (deamination product of guanine), and other base derivatives [15-20]. Family 6 enzymes, found in archaea, bacteria and eukaryotes, are known for their predominant hypoxanthine DNA glycosylase (HDG) activity [21].

As first discovered in *E. coli* back in 1974, family 1 UNG is a prototypical DNA repair glycosylase in UDG superfamily [22]. Extensive investigations in *E. coli*, human and Herpes virus UNGs have revealed its active site organization, catalytic mechanism and roles in repair of cytosine deamination and removal of uracil due to dUTP misincorporation in DNA [4, 5, 8, 23-25]. Extensive structural, mutational and biochemical investigations have revealed that family 1 UNG enzymes have evolved an elaborate mechanism for base damage recognition and glycosidic bond cleavage [4, 26]. For the uracil recognition, a pinch-push-pull mechanism is proposed, in which a proline-rich loop and a Gly-Ser loop are involved in initial *pinching*, causing the bending of the DNA backbone and facilitating the flipping of the uracil base; the Leu residue (L272 in human UNG) in the minor-groove intercalation loop (motif 2) *pushes* the uracil out of the helix by inserting into the original uracil space; the residues directly interacting with the flipped-out uracil serves the purpose of *pulling*. The chemical catalysis of the glycosidic bond cleavage requires water activation by an Asp residue (D145 in human UNG) and a His residue (H268 in human UNG) in motif 2 to act as a general acid to stabilize the leaving uracilate anion [4, 25]. Insights gained from studying family 1 UNGs have provided a framework and served as a reference for understanding other families in UDG superfamily. An interesting aspect of family 1 UNG research is the discovery of a protein

mimicry of DNA, Ugi for uracil-DNA glycosylase inhibitor, in *Bacillus* phages PBS1 and PBS2 [27-30]. Unlike regular DNA-based genomes, thymine is replaced by deoxyuracil in PBS2 genome [31]. To maintain viability in its *Bacillus* host, PBS2 genome encodes the *ugi* gene to inhibit the host UNG. Ugi is a potent inhibitor of UNG and it binds to UNG in a 1:1 stoichiometry [32, 33]. Structural analyses show that the Ugi protein forms a mimicry of DNA by inserting a phosphate backbone mimicking β-strand into the enzyme's DNA binding interface, blocking the interaction between UNG and DNA [34-37]. Ugi not only blocks the enzymatic activity of *Bacillus* UNG, it also inhibits family 1 UNG from other species including human UNG [34, 35, 38]. However, enzymes from other families in UDG superfamily are not inhibited by Ugi, suggesting a specific interaction between Ugi and family 1 UNG.

Nitratifractor salsuginis (Nsa) is nitrate-reducing chemolithoautotrophic bacterium isolated from a deep-sea hydrothermal vent chimney in the Mid-Okinawa Trough in Japan [39]. In mining sequenced genome database, we identified a family 1 UNG homolog in Nsa genome. Here, we report characterization of Nsa UNG using a combined biochemical, structural and enzyme kinetics approach. Using enzyme kinetics analysis, we found Nsa UNG is a robust uracil DNA glycosylase with narrow specificity. X-ray crystallographic analysis identified regions unique to Nsa UNG. Site-directed mutagenesis coupled with structural information allowed identification of catalytic residues common to other family 1 UNG enzymes and unique to Nsa UNG. Enzyme activity assays performed in the presence of Ugi indicated a complete lack of inhibition, a feature distinctly different from conventional family 1 UNG. The possible reasons for the

lack of Ugi inhibition are explained in light of molecular modeling analysis.

III. Experimental procedures

Cloning, site-directed mutagenesis, expression and purification of Nsa UNG-The Nsa *ung* gene from *Nitratifractor salsuginis* strain E9I37-1 (DSM 176511) (GenBank accession number: YP_004167197.1) was amplified by PCR using the forward primer NsaUNG F (5'- GGGAATTCCATATGACCAATAACAAATAC -3'; the NdeI site is underlined) and the reverse primer NsaUNG R (5'-CCGCTCGAGTCTTTT GAGCAGCAGGTC-3'; the XhoI site is underlined).The PCR reaction mixture (50 μl) consisted of 20 ng Nitratifractor salsuginis genomic DNA, 500 nM forward and reverse primers, 1 x Phusion DNA polymerase buffer, 200 μM each dNTP and 0.2 unit of Phusion DNA polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 98°C for 30 s; 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 s, annealing at 65°C for 20 s, and extension at 72°C for 40 s; and a final extension step at 72°C for 3 min. The PCR product was purified by gel DNA recovery kit (Generay). The purified PCR product and plasmid pET21a were digested by NdeI and XhoI, purified by gel DNA recovery kit, and ligated according to the manufacturer's instructional manual. The ligation mixture was transformed into *E. coli* strain HB101 competent cells by electroporation. The sequence of the Nsa UNG gene in the resulting plasmid (pET21a-NsaUNG) was confirmed by DNA sequencing.

The resulting plasmid with wild-type Nsa UNG was used as the template plasmid for all other Nsa UNG mutants. Amplification of the mutant DNA and DpnI mediated

site-directed mutagenesis procedures were modified as previously described by using primers carrying the desired mutations [40]. Briefly, PCR mixtures (25 µl) contained 10 ng of pET-21a (+)-NsaUNG as a template, 65 nM of each primer pair, 200 µM each dNTP, 1 x Phusion PCR polymerase buffer, and 1 unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation step at 98°C for 2 min; 25 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for 5 min; and a final extension step at 68°C for 10 min. After treatment with 2 units of DpnI for 1 h at 37°C, 5-µl PCR products were transformed into *E. coli* DH5α competent cells. For mutant with deleted region, we used the traditional overlapping PCR methods as described previously [41]. Successful mutant in the resultant clones were confirmed by DNA sequencing. The pET21a-NsaUNG WT and mutants were transformed into E. coli strain BH214 (ung-, mug-) by the standard protocol to express the C-terminal His-6-tagged Nsa UNG protein. Induction, sonication and purification were carried out as previously described with modification [21]. Briefly, after HisTrap column purification, fractions of the eluate containing Nsa UNG were pooled, dialyzed to buffer containing 20 mM Tris-HCl (pH 7.2), 40 mM NaCl and 1 mM DTT and loaded onto a 1-ml HiTrap SP column, washed with 5 ml of HiTrap SP buffer A (20 mM Tris-HCl (pH 7.1), and 1 mM DTT), and eluted with a linear gradient of 0–100% HiTrap SP buffer B (HiTrap SP buffer A and 1 M NaCl). Fractions containing Mba glycosylase (10–20% HiTrap SP buffer B) were pooled and concentrated through Microcon YM 10 (Millipore).

Preparation of SeMet-labeled protein-A 20-mL overnight Luria-Bertani culture of the WT NSA UNG/pET-21a (+) containing 50-μg/ml ampicillin was prepared. The next morning the cells were resuspended in 10-mL M9 medium after centrifugation and were transferred to the 1L of M9 salt medium containing 50-μg/ml ampicillin. When OD600 reached 0.5, an amino acid mixture containing 100 mg of lysine, phenylalanine, and threonine, 50 mg of isoleucine, leucine, valine, and 60 mg of SeMet was added. Then the growth temperature was reduced to 30ºC. The culture was induced with 0.5 mM IPTG and kept shaking for 16~18 hours. The cells were harvested by centrifugation, resuspended in cold nickel-nitrilotriacetic acid (Ni-NTA) buffer A containing 20 mM Trsi-HCl pH 8.0, 250 mM NaCl, 10 mM imidazole, 10 mM β -mercaptoethanol (β-ME), and 1 mM phenylmethylsulfonyl fluoride (PMSF). The cells were disrupted by ultrasonication and the supernatant was obtained by centrifugation at 14,000 rpm for 1 h at 4°C. The supernatant was then applied onto Ni-NTA affinity resin (Qiagen) equilibrated with Ni-NTA buffer A. The target protein was eluted with Ni-NTA buffer B (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 250 mM imidazole, 10 mM β-ME and 1 mM PMSF). The target protein fractions were pooled and dialyzed in a buffer consisting of 20 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 2 mM DTT. The desalted protein was applied onto a HiTrap Heparin HP column (GE Healthcare) equilibrated with Heparin HP buffer A (20 mM Tris-HCl (pH 8.0), 50 mM NaCl, 2 mM DTT), and the NSA UNG protein was eluted with 20 mM Tris-HCl (pH 8.0), 2 mM DTT with a NaCl concentration gradient from 0-1 M. The final protein was concentrated to 8 mg/ml by a Millipore centrifugal filter (molecular-weight cutoff of 10 kDa) and stored at -80°C.

Crystallization, data collection and structure determination –The initial screens for NSA UNG crystals were manually set up using the sitting-drop vapor-diffusion method, with crystal screens I and II and the index screen (Hampton Research, CA, USA). The sample was mixed with the well solution in 1:1 ratio (v/v). The SeMet-labeled wild-type protein and the native G65Y mutant were crystallized under the same condition, with the SeMet-labeled sample being kept in a reducing environment throughout. The final optimized crystallization condition is 10% PEG 3350, 0.1 M HEPES pH 7.0, 5% glycerol.

All crystals were flash frozen in liquid nitrogen after being soaked in a cryoprotectant containing all the reservoir solution components supplemented with 20% glycerol (v/v). X-ray diffraction data were collected using beamlines 17U1 (BL17U1, Y81G dataset) and 19U1 (BL19U1, WT dataset) at the Shanghai Synchrotron Radiation Facility (SSRF, Shanghai, P. R. China). The complete single-wavelength anomalous dispersion (SAD) dataset of the SeMet-labeled WT crystals was collected at the peak position of SeMet (0.979 Å). The diffraction images were processed using HKL3000 [42] and determined with the software package Crank2 [43]. One molecule of NSA UNG was predicted to be present in the asymmetric unit, with the space group P212121. After the determination of the SeMet sites, density modification was carried out until an interpretable map was obtained. The model-building and refinement were carried out based on the initial model from SAD phasing by Coot and Phenix.refinement [44, 45]. The final model containing six SeMet residues was validated by Molprobity [46]. The structure of the Y81G mutant was obtained using molecular replacement with the WT

structure as the search model. All the data collection and refinement statistics are presented in Table 3.1.

Table 3. 1. Data collection and refinement statistics

	WT-SeMet (5X3G)	Y81G (5X3H)
Data collection	SSRF BL19U1	SSRF BL17U1
Wavelength	0.9791	0.99
Space group	$P2_12_12$	$P2_1$
Cell dimensions		
a, b, c (Å)	37.87, 59.39, 99.68	38.41,114.76, 55.54
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 91.4, 90.0
Resolution (Å)	50.0-2.02 (2.09-2.02)[a]	50.0-2.5 (2.59-2.50)
R_{merge}[b] (%)	12.9 (80.6)	7.9 (26.3)
$I/\sigma_{(I)}$	25.3 (3.0)	16.5 (5.8)
Completeness (%)	100 (100)	99.2 (98.5)
Redundancy	12.7 (13.0)	3.5 (3.4)
Refinement		
Resolution (Å)	31.94-2.02 (2.18-2.02)	39.90-2. 50 (2.69-2.50)
No. reflections	15253	16470
R_{work}[c]/R_{free}[d]	0.213/ 0.261	0.211/0.265
No. atoms		
Protein	1967	3837
Water	105	24
B-factors (Å^2)		
Protein	34.57	41.42
Water	35.83	36.38
R.m.s deviations		
Bond lengths (Å)	0.003	0.006
Bond angles (º)	0.65	1.00
Ramachandran favored (%)	96.76	98.78
Allowed (%)	3.24	1.22
Outliers (%)	0	0

[a]: Values in parentheses are for the highest-resolution shell.
[b]: $R_{merge} = \Sigma |(I - < I >)|/\sigma(I)$, where I is the observed intensity.
[c]: $R_{work} = \Sigma_{hkl} ||Fo| - |Fc||/ \Sigma_{hkl} |Fo|$, calculated from working data set.
[d]: R_{free} is calculated from 5.0% of data randomly chosen and not included in refinement.

Oligodeoxynucleotide substrates –Oligodeoxynucleotides containing deoxyuridine(U), deoxyinosine (I) and deoxyxanthosine (X) were obtained or constructed as previously described [13].

DNA glycosylase activity assay –DNA glycosylase cleavage assays for Nsa UNG were performed at optimized temperature 60°C for 60 min in a 10 µl reaction mixture containing 10 nM oligonucleotide substrate, 100 nM Nsa UNG, 20 mM Tris-HCl buffer (pH 7.5), 350 mM KCl, 1mM DTT, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95°C for 5 min after adding 1 µl of 1N NaOH. The reaction mixtures (2 µl) were mixed with 7.8 µl Hi-D formamide and 0.2 µl GeneScan 500 LI Size Standard and analyzed by Applied Biosystems 3130xl genetic analyzer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

Enzyme kinetic analysis –Uracil DNA glycosylase assays were performed at optimized condition with 10 nM G/U substrates with enzyme in excess ranging from 50 nM to 2000 nM. Samples were collected at 0.01 s, 0.05 s, 0.1 sec, 0.2 s, 0.5 s, 1 s and 2 s for Nsa UNG-WT using a rapid quench flow apparatus as described below. Samples of Nsa UNG mutants were collected at 10 s, 30 s, 1 min, 2 min, 5 min and 10 min manually. The apparent rate constants for each concentration were determined by curve fitting using the integrated first-order rate equation (1):

$$P = P_{max}(1 - e^{-k_{obs}t}) \qquad (1)$$

Where P is the product yield, P_{max} is the maximal yield, t is time and k_{obs} is the apparent rate constant.

The kinetic parameters k_2 and K_m were obtained from plots of k_{obs} against the total enzyme concentration ([E_0]) using a standard hyperbolic kinetic expression with the program GraphPad 6 following the equation (2)[47]

$$k_{obs} = \frac{k_2\ [E_0]}{K_m + [E_0]} \qquad (2)$$

For some mutants with a large K_m in which K_m >> [E_0], the kinetic parameter k_2/K_m values were obtained from plots of k_{obs} against total enzyme concentration ([E_0]) using a linear regression with program GraphPad 6 following the equation (3)[48].

$$k_{obs} = \frac{k_2\ [E_0]}{K_m} \qquad (3)$$

Rapid chemical quench-flow measurements–The short time course DNA glycosylase assays as shown in Fig. 3.1 were performed at 60 °C in reaction buffer (20 mM Tris-HCl buffer (pH 7.5), 350 mM KCl, 1 mM DTT, and 1 mM EDTA) with 10 nM DNA substrates and 100 nM Nsa UNG protein using a Rapid Quench Flow (RQF-3, KinTek Corporation). Reactions were initiated by mixing equal volumes (~50 μl) of the 10 nM FAM-labeled substrate and 100 nM enzyme solution. At time points between 0 and 5 s, the reactions were quenched with 0.1 M NaOH delivered from the quench syringe.

Molecular modeling–The crystal structure of Nsa UNG-WT was used as a model for subsequent computational analysis. A structure of DNA with a flipped-out pseudouracil base was extracted from the crystal structure of family 1 human UNG-DNA complex (PDB 1EMH) [4] using the Swiss-Pdb Viewer (SPDBV) program [49]. The Nsa UNG-WT structure was superimposed upon the hUNG-DNA crystal structure (1EMH) using VMD 1.9.2 [50]. Removal of the hUNG protein coordinates resulted in a model of Nsa

UNG bound to DNA. Nsa UNG-S86A and Nsa UNG-D94A mutants complexed with DNA were constructed using the mutation tool in the Swiss-Pdb Viewer program and the "best rotamer" was chosen with the lowest clash score.

Molecular dynamics simulations –After building the initial complex structures, an explicit solvent system using the TIP3P water model was constructed in the VMD 1.9.2 using a suitably sized box. The minimum distance between any of the atoms of the solvated Nsa UNG-DNA complex and the box boundary was maintained to at least 9 Å. Potassium chloride ions were added to the system to achieve an electrically neutral system and the concentration was set to 350 mM. The CHARMM 27 all hydrogen force field for proteins [51] and nucleic acids [52] were used. Particle-mesh Ewald summation [53] was applied in the periodic boundaries condition for the efficient calculation of long-range electrostatic interactions. Energy minimization was performed by using 8000 steps to remove any unfavorable van der Waals clashes while minimally perturbing the original modeled structure. Using a Langevin barostat [54], an isothermal-isobaric ensemble (NPT) was constructed in NAMD program [55] and the system was heated gradually from 100 K to 333 K over a period of 400 ps. An integration time step of 1 fs was used in order to avoid any significant structural deformation during heating, equilibration and production runs. Coordinates were saved every 2 ps. Calculated RMSD demonstrated that the trajectories have stabilized in 5 ns for each simulation (data not shown). Thus a total of 5 ns equilibration followed by 7 ns production simulation were performed for each structural analysis. VMD 1.9.2 was used for visualization.

IV. Results

The predicted Nsa *ung* gene is 765 nucleotide long, which translates into 255 amino acids. In comparison, the prototypical *E. coli* UNG contains 229 amino acids. By SIAS analysis (http://imed.med.ucm.es/Tools/sias.html), the full length Nsa UNG shares with *E. coli* UNG 24%, human UNG 21%, human herpes simplex virus 1 (HSV1) UNG 24% sequence identity and with *E. coli* 26%, human UNG 23%, HSV1 UNG 25% sequence similarity, respectively. As shown in Fig. 3. 2A, compare to other traditional family 1 UNG enzymes, there are two major insertions in Nsa UNG labelled as region B and region C. However, like other typical family 1 UNG enzymes, Nsa UNG contains three important motifs that define the substrate specificity and catalytic activity (Fig. 3. 1A). Using fluorescently labeled DNA, we tested the DNA glycosylase activity on uracil (U)-, hypoxanthine (I)-, and xanthine (X)-containing substrates (Fig. 3.1B and C). Similar to other family 1 UNGs, Nsa UNG doesn't show any activity to hypoxanthine and xanthine-containing substrates, it only exhibited quite robust enzymatic activity on U-containing DNAs, in particular in double-stranded U-containing substrates, in which the turnover was completed in the first time point (10 min) at optimized salt condition (Fig. 3.1D, E and F). In the shorter time-course analysis carried out using a quench-flow apparatus, the reactions of double-stranded U-containing substrates were completed within five seconds (Fig. 3.1G). Similar to other family 1 UNG enzymes, Nsa UNG showed no activity towards thymine in a G/T base pair (data not shown).

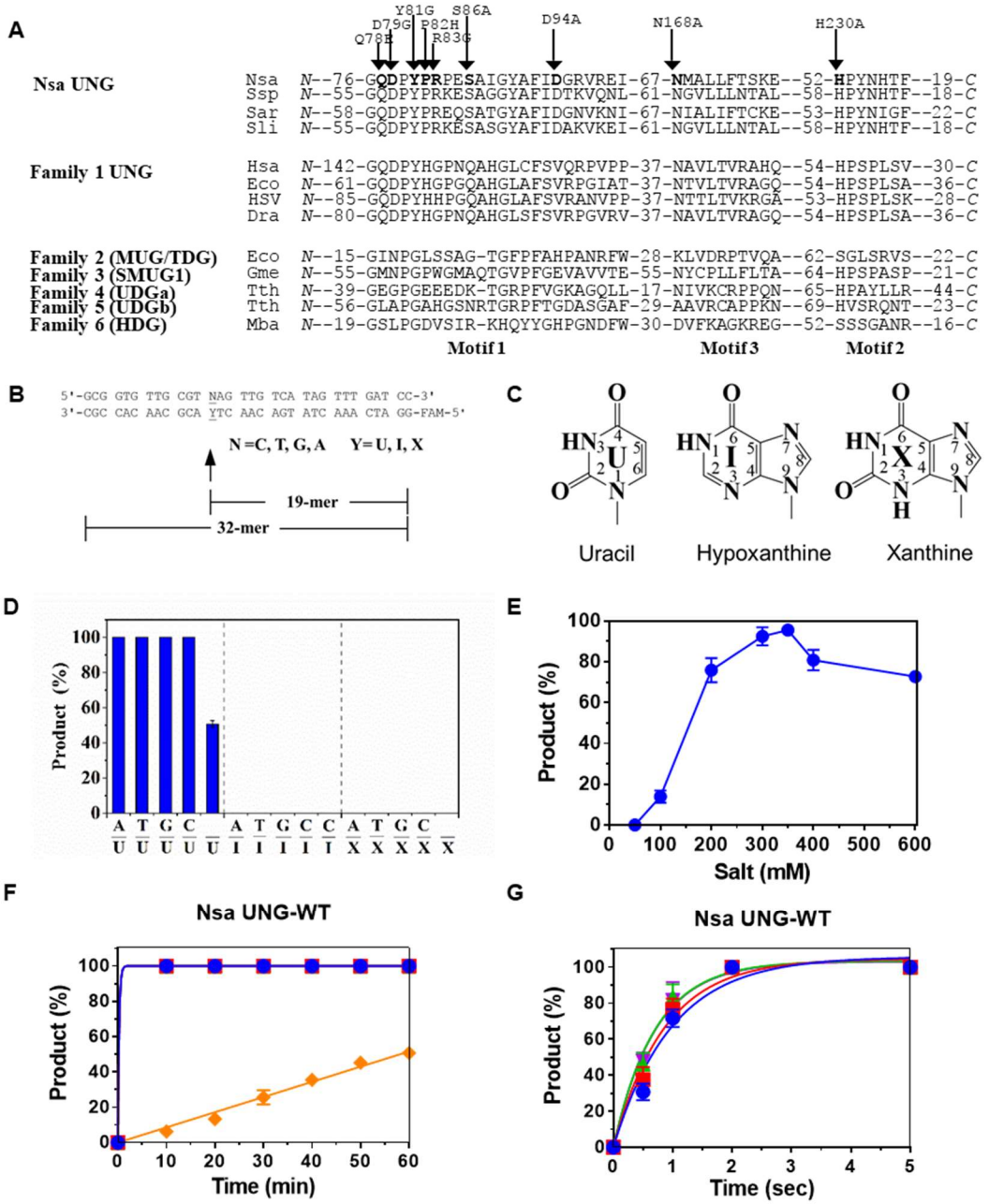


Figure 3.1. Sequence alignment, substrate and glycosylase activity of Nsa UNG. A. Sequence alignment of Nsa UNG and other UDG families; Nsa, *Nitratifractor. salsuginis*, YP_004167197.1; Ssp, *Sulfurovum sp.* AR, WP_008244665.1; Sar, *Sulfurospirillum arcachonense*, WP_024954916.1; Sli, *Sulfurovum lithotrophicum*, WP_046550169.1; Family 1 (UNG): Hsa, *H. sapiens*, NP_003353; Eco, *E. coli*, NP_289138; HSV, *Herpes Simplex Virus 1*, P10186.1. Dra, *Deinococcus radiodurans* R1, NP_294412; Family 2 (MUG/TDG): Eco, *E. coli*, P0A9H1; Family 3 (SMUG1):

Gme, *G. metallireducens* GS-15, YP_383069; Family 4 (UDGa): Tth, *Thermus thermophilus* HB27, YP_004341.1; Family 5 (UDGb): Tth, *T. thermophilus* HB8, YP_144415.1; Family 6 (HDG): *Methanosarcina* barkeri str. Fusaro, YP_304295. B. Sequence of uracil, hypoxanthine, xanthine-containing DNA substrates. C. Chemical structures of uracil, hypoxanthine and xanthine. D. DNA glycosylase activity of WT Nsa UNG on U-, X-, I-containing substrates. Cleavage reactions were performed as described in Material and Methods with 100 nM WT Nsa UNG and 10 nM substrate and incubated for 1 hour. Data are the averages of three independent experiments. E. Effect of salt to the activity of Nsa UNG. The UDG activity assays were performed with 5 uM NsaUNG and 0.5 uM G/U substrate in presence of different concentration of KCl, samples were quenched at 0.2 s. F. Time course analysis of DNA glycosylase activity of WT Nsa UNG on U-containing DNA substrates. (●) A/U, (■) T/U, (▲) G/U, (▼) C/U, (♦) single-stranded U. The assay was performed as described in Experimental Procedures under DNA glycosylase activity assay. G. Short time course analysis of DNA glycosylase activity of WT Nsa UNG on double-stranded U-containing DNA substrates. (●) A/U, (■) T/U, (▲) G/U, (▼) C/U. The assay was performed as described in Experimental Procedures under DNA glycosylase activity assay and the reactions were quenched at specific time points as indicated.

Even though Nsa UNG contains three catalytically important motifs, the sequences within the motifs show some distinct differences between Nsa UNG and *E. coli* UNG. Moreover, in comparison with *E. coli*, human and herpes simplex virus-1 (HSV-1) UNGs, Nsa UNG possesses insertions (region B and region C) and deletions (region E) in several regions outside of the catalytic motifs (Fig. 3.2A). To understand the structural differences between Nsa UNG and other previously studied UNG enzymes, we solved crystal structures of the WT enzyme (PDB 5X3G) and the Y81G mutant (PDB 5X3H). Initial attempts to solve the enzyme structure using molecular replacement were not successful. Therefore, the single-wavelength anomalous dispersion (SAD) method was employed to solve the WT structure, using selenomethionine as the anomalous scatterer. The finished model of the wild type enzyme contains 249 residues (T7-R255). The space group belongs to the $P2_12_12$ space group and the SeMet crystal diffracted to a high resolution of 2.02 Å with 99.4% completeness (Table 3.1). The final model also

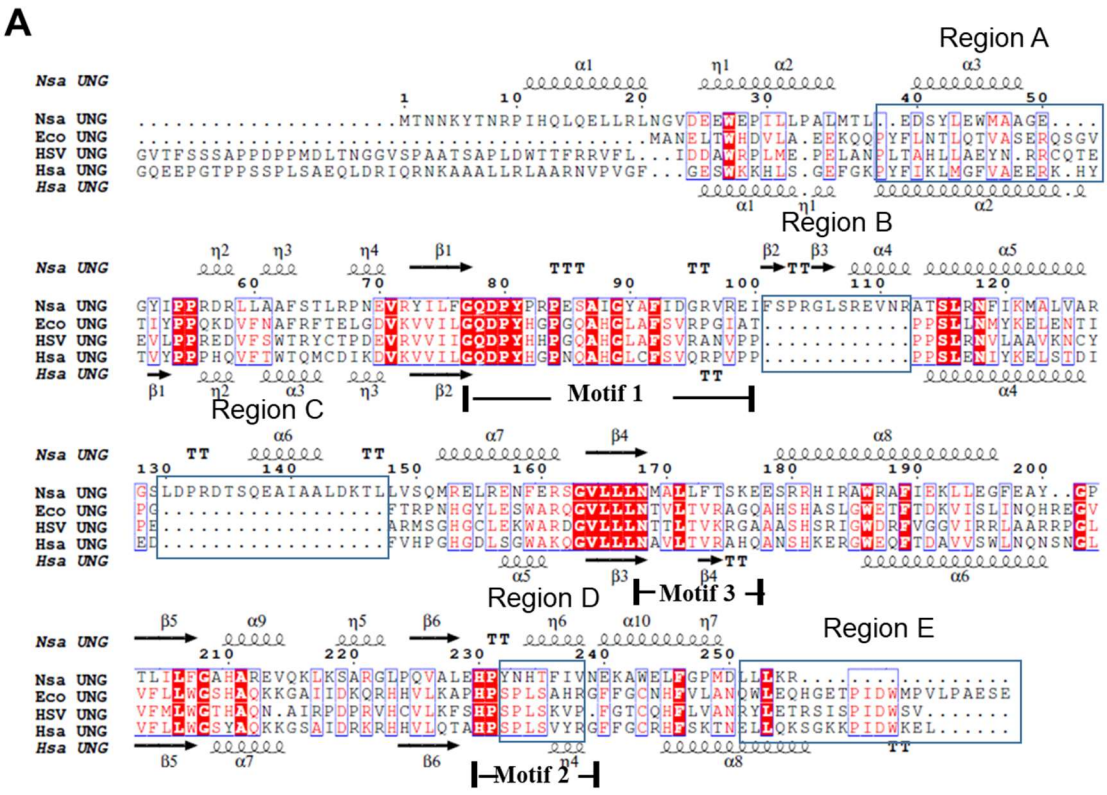
contains 111 water molecules. All the residues are in good geometry with no residues falling in the Ramachandran outlier region. The wild type Nsa UNG structure showed a central parallel four-strand β-sheet surrounded by α-helices on each side (Fig. 3.2B, C and D).

A DALI search revealed several homologs that shared structural similarity with Nsa UNG, and the top three homologs are HSV-1 uracil DNA glycosylase in complex with an inhibitor protein P56 (PDB 4L5N), uracil DNA glycosylase from cod in complex with the proteinaceous inhibitor Ugi (PDB 4LYL), and uracil DNA glycosylase from herpes simplex virus 1 in the apo form (PDB 1LAU). The RMSD is 2.45 Å over 188 Cα atoms between Nsa UNG and the closest structural homolog HSV-1 UNG. Although the general fold of the enzymes is well preserved, quite a few regions in Nsa UNG display relatively large structural variations. In region A, the deletion of four amino acids resulted in a shorter α-helix in Nsa UNG (Fig. 3.2A and E). In region B, the 12-residue insertion encompassing R98-A113 formed a longer loop and an extra helix (α4) (Fig. 3.2A and E). In region C, the 18-residue insertion covering G128-S150 formed a long loop with the extra α6 helix in the middle (Fig. 3.2A and E). Also, in motif 2, the Nsa UNG has a longer helix but shorter loop compared to human UNG (Region D). On the other hand, the C-terminus of Nsa UNG (Region E) is much shorter, as compared to the loops formed in other UDGs. Additionally, the axes of several helices of Nsa UNG (α2, α3, α8 and α10) point to different directions other than the consensus directions displayed by its homologs. These local structural variations together contribute to the relative large RMSD value and explained the failure to obtain the solution by molecular replacement.

The structural and sequence alignments clearly indicate distinct differences between Nsa UNG and conventional family 1 UNG. To understand how these differences impact the glycosylase function, we conducted a series of mutational and enzyme kinetics analyses. We initially deleted the extra region B and region C separately to test the effect on glycosylase activity. Neither deletion B (del101-112) nor deletion C (del130-147) showed any glycosylase activity. These results suggest that although region B and region C are not found in conventional family 1, they are essential for Nsa UNG's enzymatic function. The extra secondary structures in regions B and C may be needed for the integrity of Nsa UNG's structure. To examine this possibility, we took a closer look at regions B and C. Spatially, these two regions are next to each other (Fig. 3.2E). Interestingly, the side chain of R96 in region B forms a 2.0-Å salt bridge with E109 of α4; the side chain of Q137 in region C forms a 2.9-Å hydrogen bond with the main chain of L106; and the side chain of F101 in region B stacks onto the side chain of K145 in region C (Fig. 3.2F). These inter-regions interactions may stabilize the Nsa UNG's structure.

The catalytic mechanism in family 1 UNG is extensively investigated through biochemical, structural and chemical studies [25]. The general theme is that a uracil base is contacted by the main chain of the Gln residue in motif 1 (Q in GQDP, Q78 in Nsa UNG), the side chain of the first His residue in motif 2 (H230 in Nsa UNG) and the side chain of the first Asn in motif 3 (N168 in Nsa UNG), in which the His residue serves as a general acid to promote the departure of the uracilate anion leaving group to form an oxocarbenium intermediate in a S_N1 mechanism. The Asp residue in motif 1 (D in

GQDP, D79 in Nsa UNG) acts as a general base to activate a water molecule for the hydrolysis of the N-glycosidic bond. These four amino acid residues are conserved in Nsa UNG and spatially are located in the similar orientation towards uracil (Fig. 3.2D). We measured kinetic constants of Q78E, D79G, N168A and H230A using G/U substrate. Because the catalytic activity in some of the mutants was much reduced, which did not allow determination of kinetics using conventional steady-state approach, we determined K_m and k_2 values using the methods as previously described (Table 3.2 and Fig. 3.3) [41, 47, 48]. In some cases, the individual K_m or k_2 values of the mutants could not be determined due to a very high K_m that did not allow for saturation of the substrate (a straight line in Fig. 3.3B). Q78E did not retain any glycosylase activity (Table 3.2). D79G had a relatively small effect on K_m but the k_2 value was reduced by over seventeen-hundred fold, resulting in a seventy-five-hundred-fold difference in k_2/K_m (Table 3.2). On the other hand, N168, which interacts with N3 and O4 of uracil, lost substantial binding affinity to the substrate by an alanine substitution (N168A), resulting in a k_2/ K_m that was a million-fold lower than the wild type enzyme (Table 3.2). H230A mutation increased the K_m value by thirty-two-fold and reduced the k_2 value by close to six-thousand-fold (Table 3.2). This is consistent with the catalytic role played by these residues in family 1 UNG enzymes.


A
Region A
Region B
Region C
Region D
Region E
Motif 1
Motif 2
Motif 3
Nsa UNG
Eco UNG
HSV UNG
Hsa UNG
Hsa UNG

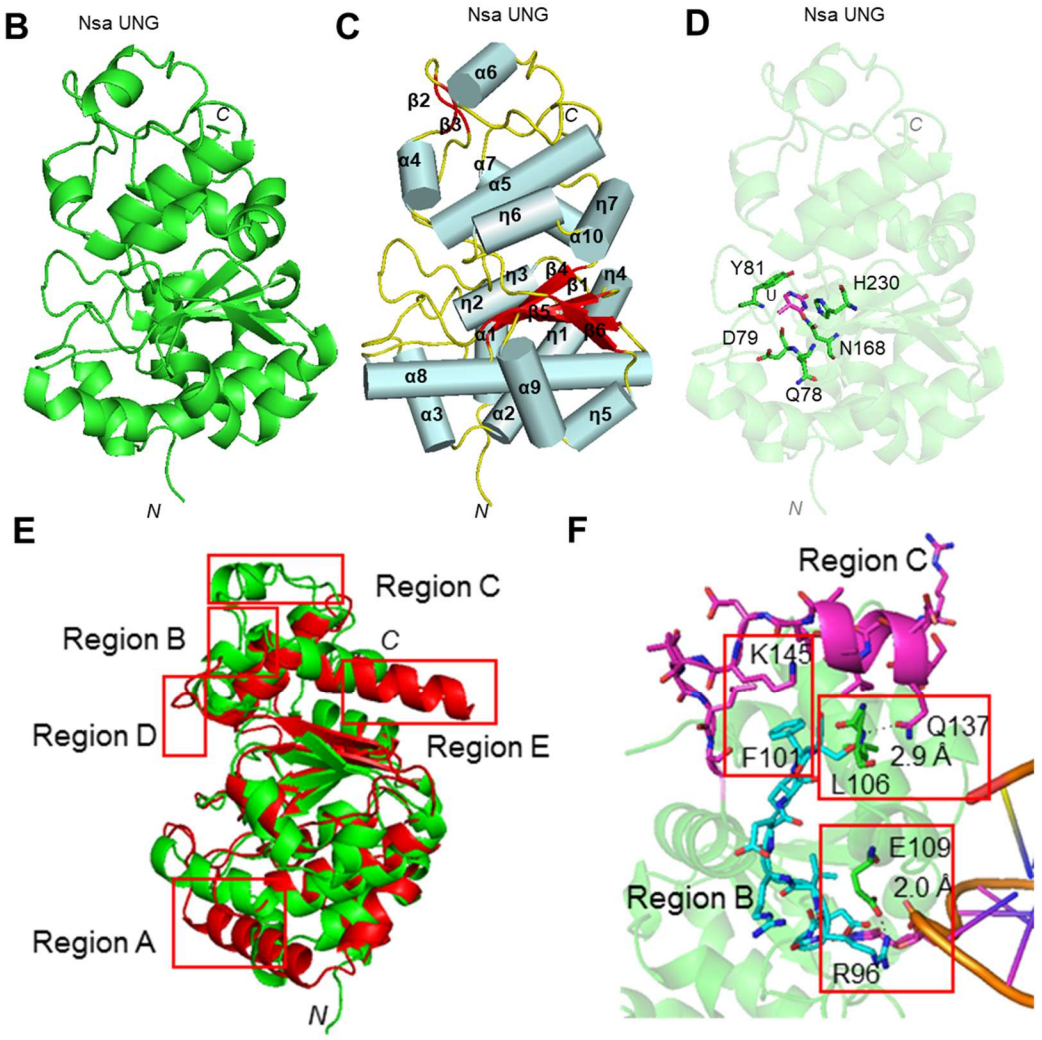


Figure 3. 2 Sequence alignment and structural comparison of Nsa UNG and human UNG. Structural diagrams were constructed using PyMOL. A. Secondary structure-based sequence alignment of Nsa UNG and other family 1 UNG enzymes. Nsa UNG: Nsa, *Nitratifractor. salsuginis*, YP_004167197.1; Eco UNG: Eco, *E. coli*, NP_289138; HSV UNG, *Human Herpes Simplex Virus* 1, P10186.1; Hsa UNG, Hsa, *H. sapiens*, NP_003353. The alignment was constructed using ESPript 3.0 [56]. B. Overall structure of Nsa UNG shown in cartoon mode. C. Secondary structures of Nsa UNG. Helices are shown as blue column and beta strands are shown in red arrow. D. Active site of Nsa UNG. Nsa UNG structure is shown in cartoon mode and catalytic residues (Tyr 81, Asp 79, Gln 78, Asn168 and His230) are shown in licorice. E. Superimposition of Nsa UNG (in green) with human UNG (PDB: 1EMH, in red). The secondary structural differences are boxed in red, corresponding to structure-based sequence alignment shown in Fig. 3. 2A. F. Close-up view of interactions between Region B (in cyan) and Region C (in purple) in Nsa UNG. Residues involved in interactions are boxed in red.

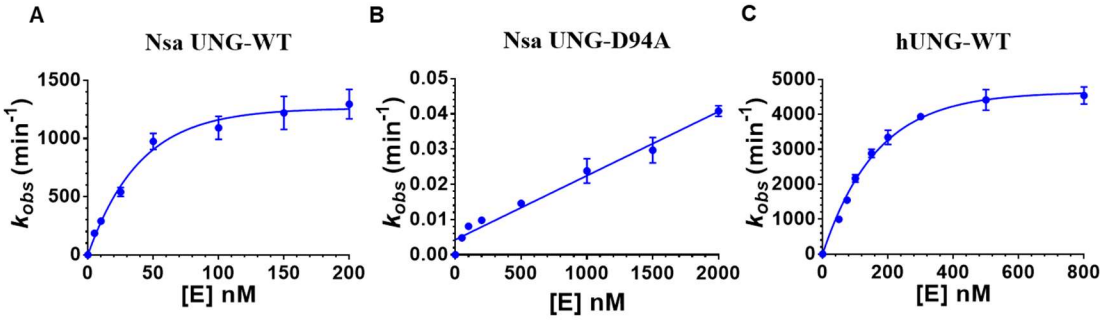


Figure 3. 3 Representative Kinetic analysis of Nsa UNG-WT, Nsa UNG-D94A and hUNG. See Experimental Procedures under enzyme kinetic analysis for details. A. Nsa. UNG-WT. B. Nsa UNG-D94A. C. hUNG-WT. The assay for hUNG was performed similar to Nsa UNG except that the KCl concentration was 30 mM.

Table 3. 2. Kinetic constants of Nsa UNG on G/U substrate[a]

Enzymes	K_m (nM)	k_2 (min^{-1})	k_2/K_m (min^{-1} nM^{-1})
wild type	27 ± 2.9	1300 ± 120	48
Q78E	N. A.[b]	N. A.	N. A.
D79G	118 ± 15	0.76 ± 0.05	6.4 x 10^{-3}
Y81G	N.A.	N.A.	N.A.
R83G	N.D.[c]	N.D.	4.0 x 10^{-5}
P82H	320 ± 45	4.3 ± 0.23	1.3 x 10^{-2}
S86A	101 ± 8.8	3.5 ± 0.07	3.4 x 10^{-2}
D94A	N. D.	N.D.	1.2 x 10^{-3}
N168A	N. D.	N.D.	4.0 x 10^{-5}
H230A	852 ± 106	0.22 ± 0.03	2.5 x 10^{-4}

[a]: The reactions were performed as described in Experimental Procedures under enzyme kinetic analysis. Data are showed as average ± SD from three independent experiments. .

[b]: No activity detected under the assay conditions.

[c]: Not determined. Individual K_m and k_2 values were not determined due to a relative large K_m.

In addition to the four invariant catalytic residues, motif 1 in Nsa UNG also contains several other conserved residues (Fig. 3.1A). Given the role of motif 1 in protein-DNA interactions, we generated a model of Nsa UNG-DNA by superimposing apo Nsa UNG onto DNA-bound hUNG (PDB 1EMH) (Fig. 3.4). For comparison, the complex structure of hUNG is shown in Fig. 3.4B and the superimposition of the two is shown in Fig. 3.4C. A close examination of the protein-DNA interface revealed that the side chain of R83 was located in close proximity to the DNA phosphate backbone, allowing a direct interaction with the phosphate 5' to the deoxyuridine (Fig. 3.4A and D). More interestingly, the side chain of R83 also formed a salt bridge with the invariant D94 (Fig. 3.4D). The side chain of D94, in turn, was stabilized by a hydrogen bond with the side chain of S86 (Fig. 3.4D). In comparison, the corresponding positions in hUNG are occupied by G149, S159 and Q152, respectively (Fig. 3.4B). The lack of a functional side chain in G149 does not allow any interaction with the DNA backbone. Accordingly, we tested the mutational effects of P82H, R83G, D94A and S86A on the glycosylase activity. P82 is conserved in this group of UNG enzymes (Fig. 3.1A), Substitution of P82 with histidine, the residue commonly found in conventional family 1 UNGs, caused more than 10-fold increase in K_m and close to 300-fold reduction in k_2 (Table 3.2). Evidently, both R83G and D94A affected the K_m substantially (Table 3.2). S86A lowered the binding affinity by less than 4-fold as judged by K_m values, which is consistent with its supporting role to the salt bridge. While S86A and D94A reduced the k_2/K_m value by three and four orders of magnitude, respectively, R83G decreased it by six orders of magnitude (Table 3.2). This is consistent with the role of R83, which directly interact

with the backbone. Both the sequence and structural comparisons suggest that the salt bridge network is a unique feature of Nsa UNG enzyme. The extra positive charge and the R83-mediated salt bridge are consistent with a K_m of 27 nM in Nsa UNG as compared with a K_m of 169 nM in hUNG (Table 3.2 and Fig. 3.3). To better understand the role of the networked interactions on R83-mediated DNA backbone interactions, we performed molecular dynamics simulations of the S86A and D94A mutants along with the Nsa UNG-WT (Fig. 3.4E, F and G). Consistent with the kinetic analysis, the R83-DNA interaction had the highest probability at 2.7 Å for the S86A mutant as compared with the highest probability at 2.6 Å for the wild type enzyme (Fig. 3.4E and F). However, the R83-DNA interaction had the widest distribution with an average distance of 4.2 Å for the D94A mutant, resulting in a dramatic change in binding affinity and catalytic activity (Fig. 3.4E and G). We also performed molecular dynamics simulations of the P82H to understand its effect on the salt bridge. Once P82 is substituted by His residue, R83 is no longer consistently kept in close proximity to the phosphate backbone (Figs. 3.4H, I, J), which results in three orders of magnitude reduction in k_2/K_m value (Table 3.2). These data indicate an evolution of an elaborate salt bridge network.

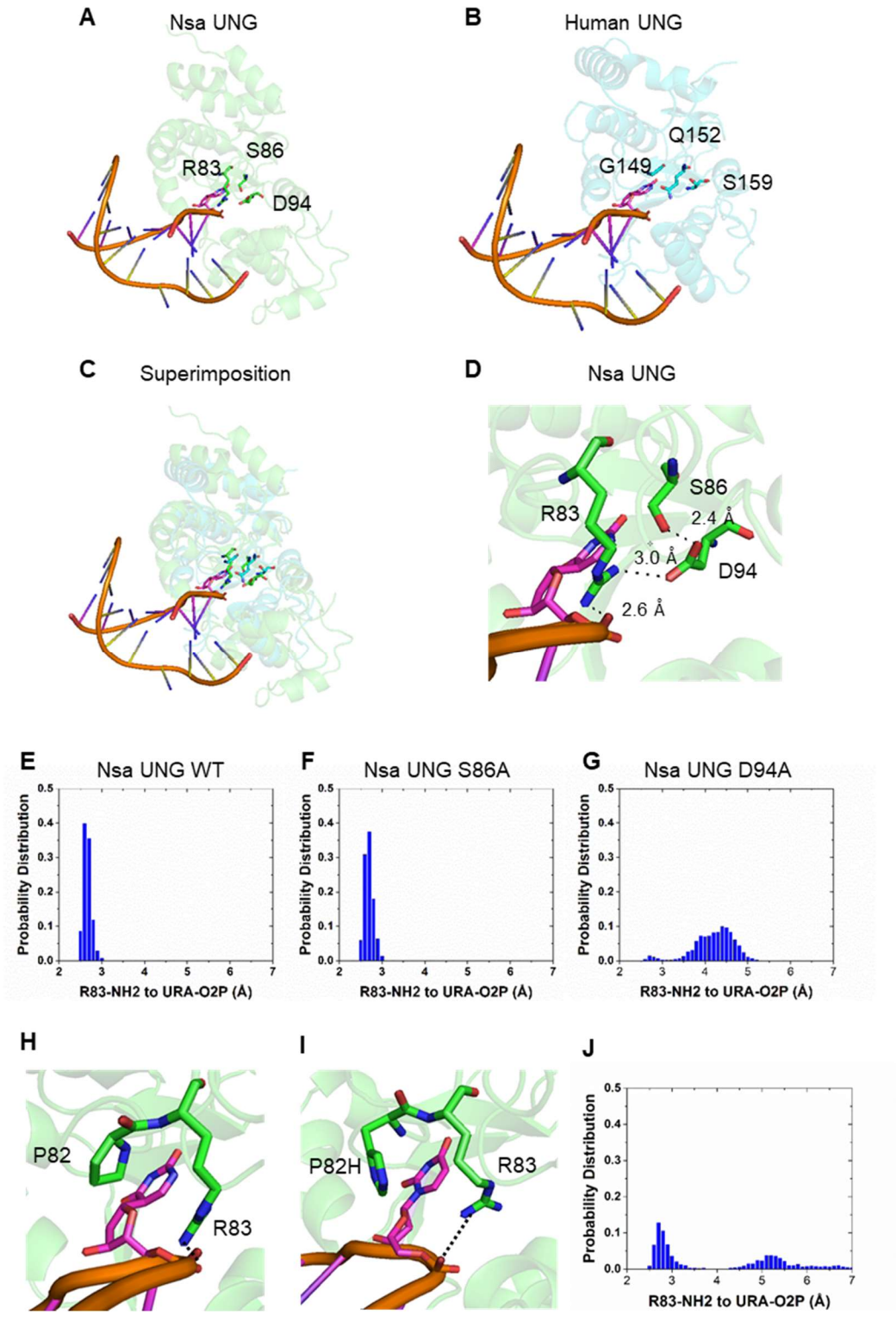

A
Nsa UNG
S86
R83
D94
B
Human UNG
Q152
G149
S159
C
Superimposition
D
Nsa UNG
S86
R83
2.4 Å
3.0 Å
D94
2.6 Å
E
Nsa UNG WT
F
Nsa UNG S86A
G
Nsa UNG D94A
Probability Distribution
R83-NH2 to URA-O2P (Å)
H
P82
R83
I
P82H
R83
J
Probability Distribution
R83-NH2 to URA-O2P (Å)

Figure 3. 4 Comparison of protein-DNA interactions between Nsa UNG and human UNG. A. The interactions between Nsa UNG-R83 and modeled uracil-containing DNA (DNA model was taken from PDB 1EMH). B. The interactions between human UNG and Uracil-containing DNA (PDB 1EMH). C. Superimposition of the interactions between modeled Nsa UNG-DNA complex and human UNG-DNA complex. D. Close-up view of the networked salt-bridge in Nsa UNG that involves in DNA phosphate backbone interaction. The network of side chain interactions among R83, S86 and D94 are shown in licorice. Probability distribution of distance between R83 and DNA backbone in Nsa UNG-WT (E), Nsa UNG-S86A (F). Nsa UNG-D94A (G). Consequence changes of the interaction between R83 and DNA backbone in Nsa UNG-WT and Nsa UNG-P82H (H, I). J. Probability distribution of distance between R83 and DNA backbone in Nsa UNG-P82H

Family 1 UNG differs from other families in having a Tyr residue in GQDPY of motif 1 (Fig. 3.1A). Other families contain a Gly residue in this position. We changed Y81 to Gly and found Y81G was catalytically inactive against uracil-containing substrate or thymine in a G/T base pair. This result suggests that even though a Gly residue exists in other families, it was not compatible with family 1 UNG enzymes. To understand the structural impact of Y81G that may influence its enzymatic function, we solved the crystal structure of apo-Y81G protein, by molecular replacement using the WT structure as the search probe (Fig. 3.5A). The Y81G protein was solved at a lower resolution, but the structure is essentially the same (Fig. 3.5B). However, its space group changes to $P2_1$ with two molecules present in the asymmetric unit, although the crystals were grown under an identical condition. The RMSD between WT and Y81G is 0.3-0.4 Å over 246 aligned Cα atoms. Therefore except for slight rearrangements for a couple of 3_{10} helices (η3 and η6), the Y81G point mutation causes very few structural changes. We have encountered a similar scenario during our structural characterization on Phe SMUG2. When we introduced the G65Y mutation, the resulting structure of the mutant is highly similar to that of the WT enzyme [57].

The overall structure of Nsa UNG-Y81G is shown in Fig. 3.5A and the superimposition with the Nsa UNG-WT is shown in Fig. 3.5B, highlighting the difference in Y81 and Y81G. Overall, the two structures deviate little in the backbone arrangement. We then closely compared the subtle differences between the two structures. The first notable difference is that the Y81 forms a hydrogen bond with the main chain of F92, whereas the Y81G loses this interaction (Fig. 3.5C). The second difference is that the side chain of R83 in Nsa UNG-Y81G is invisible, suggesting that it does not form a stable salt bridge with D94 (Fig. 3.5C). The final difference is that the E109, which forms a salt bridge with R96 in Nsa UNG-WT, is oriented differently in Nsa UNG-Y81G. As a result, the salt bridge is disrupted, which may affect protein-DNA interaction or structural stability (Fig. 3.5C). These subtle changes associated with Y81G substitutions may lead to the loss of glycosylase activity. Because a Tyr in this position is unique to family 1 UNG enzymes, we subsequently examined the likely hydrogen bonding between the Tyr and the adjacent, highly conserved Phe residue. Indeed, human, *E. coli* and HSV UNG all contain a hydrogen bond (Fig. 3.6). The existence of the Tyr residue defines the narrow base recognition pocket, in which the access of thymine with a C5-methyl group is blocked by the bulky Tyr side chain. In human UNG, Y147A mutation causes over one-thousand-fold reduction in UDG activity but gains a low level activity on removal of a thymine base [58]. The hydrogen bond between the side chain of Y147 and the main chain of F158 may provide a strong interaction to enforce the rigidity of the base recognition pocket. In the case of Nsa UNG, the effect of Y81G is profound as the mutation causes loss of multiple interactions.

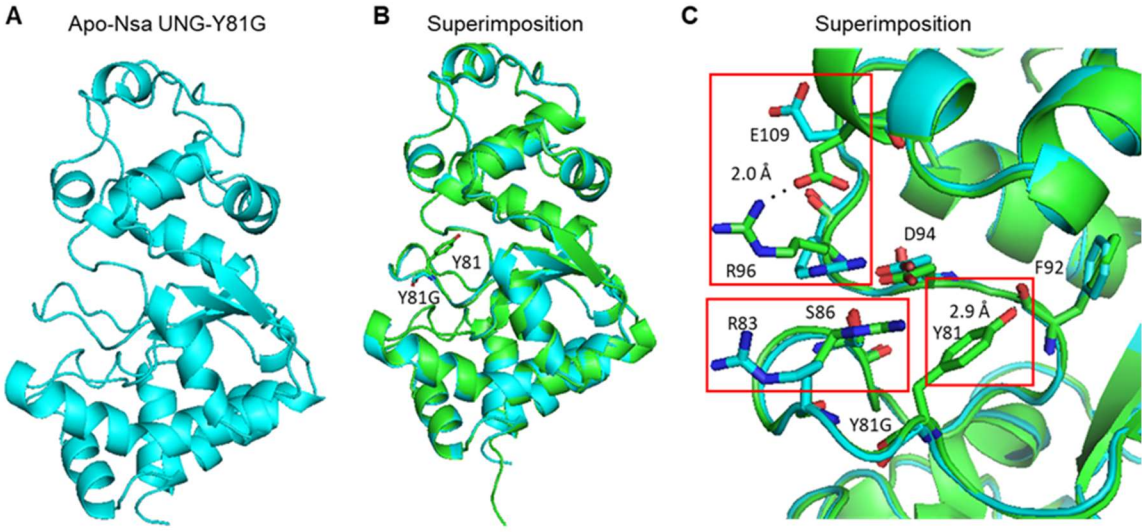


Figure 3. 5 Structure of Nsa UNG-Y81G and its comparison with Nsa UNG-WT. A. The overall structure of Nsa UNG-Y81G. B. Superimposition of Nsa UNG-Y81G with Nsa UNG-WT. C. Close-up view of the local structural differences between Nsa UNG-Y81G and Nsa UNG-WT. The differences shown in red box are highlighted in licorice.

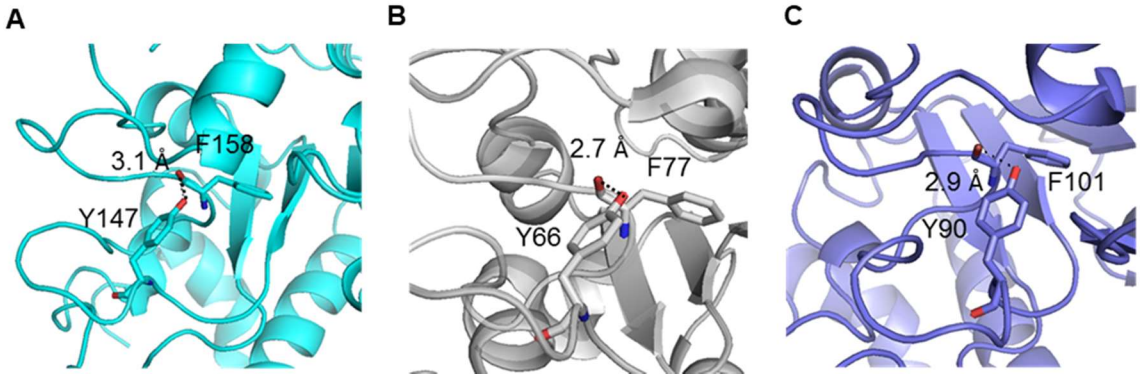


Figure 3. 6. Interaction of Tyr residue with Phe residue in family 1 UNG enzymes. A. Human UNG (PDB 1UGH). B. E. coli UNG (PDB 1LQG). C. HSV UNG (PDB 1UDI). Protein structures are shown as cartoon, Tyr and Phe residues are shown as licorice.

As mentioned above, Ugi exclusively inhibits the glycosylase activity of family 1 UNG enzymes. Given the structural differences observed in Nsa UNG protein, we set out to test the potential inhibition by Ugi. To our surprise, Nsa UNG was not inhibited by incorporation of Ugi in the glycosylase assay in a 1:1 ratio, while hUNG was inhibited

completely (Fig. 3.7A). We then increased the Ugi concentration up to 1:1000 ratio but still did not observe any indication of inhibition (Fig. 3.7A). To further confirm the lack of interactions between Nsa UNG and Ugi, we performed gel mobility analysis under non-denaturing conditions. As expected, hUNG was converted hUNG-Ugi complex while there was no indication that Nsa UNG formed a complex with Ugi (Fig. 3.7B). The potential structural explanation for complete lack of Ugi inhibition will be discussed later.

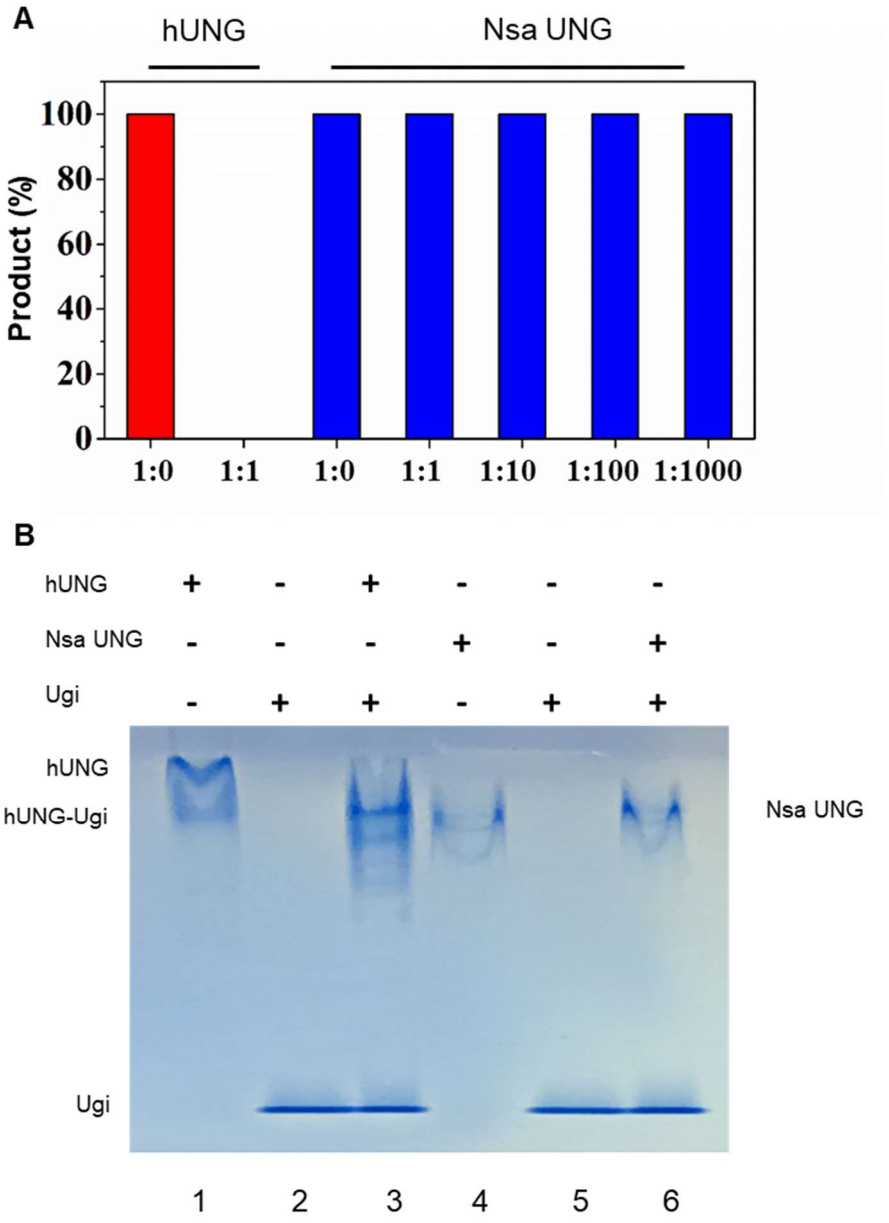


Figure 3. 7 Ugi inhibition and binding analysis. A. The reactions were performed as described under DNA golycosylase assay in Experimental Procedures with addition of indicated amount of Ugi protein. The concentrations of Nsa UNG versus Ugi are shown in molar ratio. B. Gel mobility shift analysis of binding between Ugi and hUNG/Nsa UNG[59]. After incubation of enzyme with/without Ugi, reaction products were separated on 15% native PAGE gel (pH 8.8). Lane 1: human UNG (0.1 nmol); Lane 2: Ugi (1 nmol); lane 3: human UNG: Ugi = 1:10; Lane 4: Nsa UNG (0.1 nmol); Lane 5: Ugi (1 nmol); lane 6: Nsa UNG: Ugi = 1:10.

V. Discussion

Uracil DNA glycosylase is ubiquitous in nature with homologs distributed in all three domains of life, bacteria, archaea and eukaryotes. UDG superfamily contains six families with different specificity and catalytic mechanisms. Family 1 UNG enzymes found in bacteria and eukaryotes are the most extensively studied. The readily available information from sequenced genomes provides a rich source to investigate the diversity of the UDG universe. Nsa UNG occupies a unique niche in UDG superfamily, even though the distribution of this type of family 1 UNG appears to be limited. Among the sequenced genomes available so far, we only found three Nsa UNG homologs in *Sulfurovum sp.* NBC37-1, *Sulfurospirillum arcachonense*, *Sulfurovum lithotrophicum* isolated from a deep-sea hydrothermal vent chimney in the Mid-Okinawa Trough in Japan. *Nitratifractor*, *Sulfurovum* and *Sulfurospirillum* are closely related as they all belong to Class of epsilonproteobacteria. Among the four homologs, Nsa UNG shares 48-60% sequence identity with the other three homologs. Additionally, all four homologs contain the insertions seen in regions B and C, suggesting that they share high degrees of global sequence and structural conservation that differentiate them from conventional family 1 UNG enzymes.

Besides the global structural differences revealed by sequence and structural comparison, Nsa UNG presents two special biochemical features that distinguishes it from conventional family 1 UNG. R83 provides a novel ionic interaction with the phosphate of the uridine. The break of the phosphate backbone interaction reduces the k_2/K_m by six orders of magnitude, on par with the well-known catalytic H230 in motif 2

(Table 3.2). These results underscore the critical role that the nonspecific interaction plays in maintaining Nsa UNG's glycosylase activity. Nature appears to have evolved an elegant network of interactions to support this important backbone interaction by having R83 sandwiched by two rigidifying Pro residues and stabilized by a salt bridge with the negatively charged D94, and by forming a strong hydrogen bond between S86 and D94 (Fig. 3.4). The evolution of the networked salt bridge may be related to *N. salsuginis*' adaption to its deep-sea living environment. It is well known that salt can weaken protein-DNA interactions. It is also known that ionic interactions help stabilize halophilic proteins [60]. The networked salt bridge seen in Nsa UNG may be needed to stabilize its structure and facilitate its interaction with DNA.

The second unique feature of Nsa UNG is its complete lack of inhibition by Ugi. To understand the possible structural differences accounted for the effect on Ugi inhibition, we compared the Ugi interaction interfaces of Nsa UNG and hUNG (Fig. 3.8A, B and C). Notably, in the region highlighted in a yellow box, the extended loop in Nsa UNG may collide with the Ugi, while the shorter loop in hUNG ensures surface complementarity (Fig. 3.8A, B and C). On the other hand, in the region highlighted in a red box, that hUNG has a relatively longer loop protruding into the concave surface of Ugi while the shorter loop in Nsa UNG may prevent its proper interactions with Ugi (Fig. 3. 8A, B, D, E and F). As a consequence, Ugi fails to inhibit Nsa UNG as it does for conventional family 1 UNG enzymes. To further explain the lack of Ugi inhibition, molecular dynamic simulations was performed and the free binding energy between Nsa UNG/hUNG and Ugi was calculated using the MM/PBSA method [61]. The results

showed the binding free energy (4.13 kcal/mol) between Nsa UNG and Ugi is much larger than that (-54.91 ± 1.98 kcal/mol) between hUNG and Ugi that might explain why Nsa UNG fails to interact with Ugi.

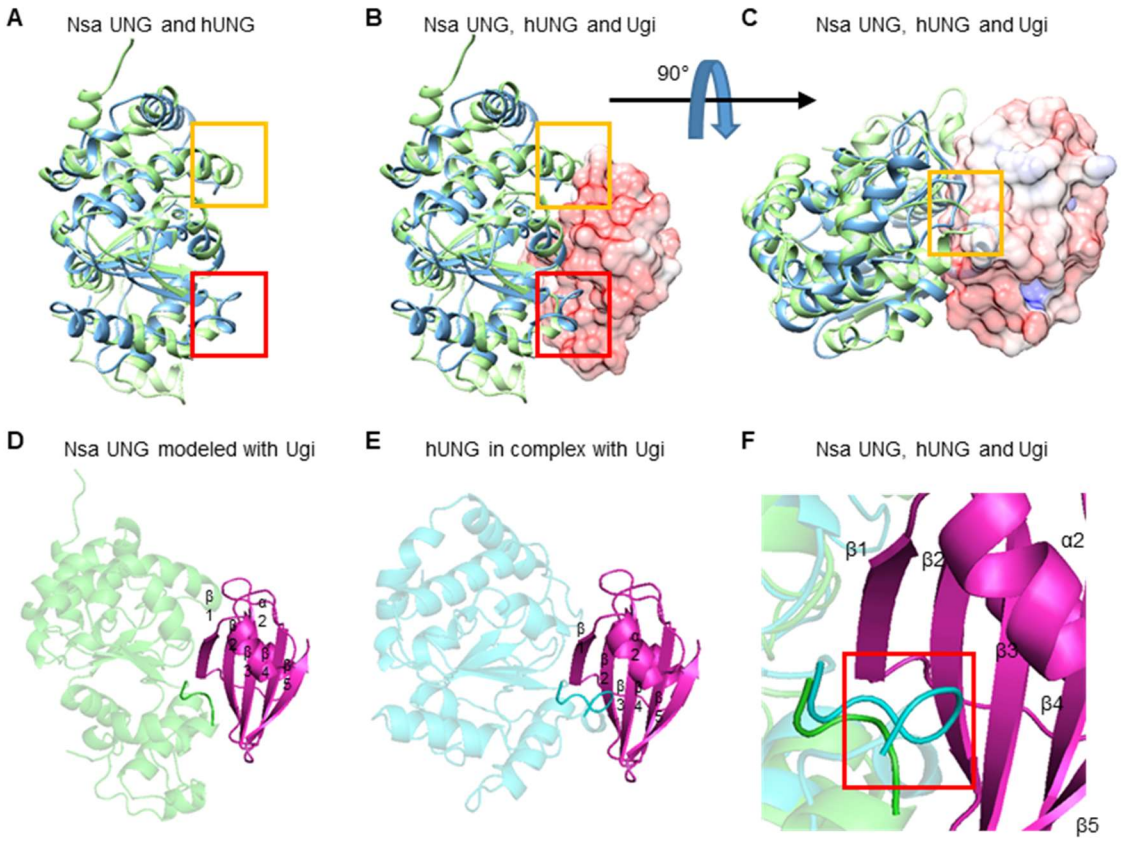


Figure 3. 8 Potential interactions between Nsa UNG and Ugi and its comparison with human UNG. A. Superimposition between Nsa UNG and hUNG. B. Superimposition between Nsa UNG and human UNG in the presence of Ugi (Ugi was taken from PDB 1UGH). C. Horizontal 90 degree rotated of B to more clearly show the clash between Nsa UNG and Ugi. The electrostatic potential was calculated by Delphi 7.0 and visualized using UCSF Chimera. The positive potential region is colored in blue and the negative potential region is colored in red. The major differences of the Ugi interaction interface between Nsa UNG and hUNG are highlighted in yellow and red boxes. D. Modeled interactions between Nsa UNG and Ugi (Ugi was taken from PDB 1UGH). E. Interactions between human UNG and Ugi (PDB 1EMH and 1UGH). F. Close-up view of the superimposition between Nsa UNG and human UNG in the presence of Ugi.

In summary, this study presents an unconventional family 1 UNG from N. *salsuginis*. Despite of the conservation of several key catalytic residues, Nsa UNG exhibits certain unique structural properties to interact with the DNA backbone. The evolution of a strong and well-positioned ionic interaction with the phosphate backbone is unseen in any other glycosylases in the UDG superfamily. More functional adaptation may be seen as we continue to explore the sequence and structural space in UDG superfamily.

VI. References

1. Lindahl, T. (1993) Instability and decay of the primary structure of DNA, *Nature.* **362**, 709-15.

2. Shapiro, R. (1981) Damage to DNA caused by hydrolysis in *Chromosome Damage and Repair* (Seeberg, E. & Kleppe, K., eds) pp. 3-18, Plenum Press, New York.

3. Suzuki, T., Yamaoka, R., Nishi, M., Ide, H. & Makino, K. (1996) Isolation and characterization of a novel product, 2'-deoxyoxanosine, from 2'-deoxyguanosine, oligodeoxynucleotide and calf thymus DNA treated by nitrous-acid and nitric-oxide, *J Am Chem Soc.* **118**, 2515-2516.

4. Parikh, S. S., Putnam, C. D. & Tainer, J. A. (2000) Lessons learned from structural results on uracil-DNA glycosylase, *Mutat Res.* **460**, 183-99.

5. Krokan, H. E., Drablos, F. & Slupphaug, G. (2002) Uracil in DNA--occurrence, consequences and repair, *Oncogene.* **21**, 8935-48.

6. Cortazar, D., Kunz, C., Saito, Y., Steinacher, R. & Schar, P. (2007) The enigmatic thymine DNA glycosylase, *DNA repair.* **6**, 489-504.

7. Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J. & Schar, P. (2001) Thymine DNA glycosylase, *Prog Nucleic Acid Res Mol Biol.* **68**, 235-53.

8. Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily, *Mutation research.* **460**, 165-81.

9. Dong, L., Mi, R., Glass, R. A., Barry, J. N. & Cao, W. (2008) Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase, *DNA repair.* **7**, 1962-72.

10. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C. & Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, *Science.* **333**, 1303-7.

11. Maiti, A. & Drohat, A. C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites, *J Biol Chem.* **286**, 35334-8.

12. Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W. & Verdine, G. L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors, *Current biology : CB.* **9**, 174-85.

13. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N. & Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG, *Journal of molecular biology.* **385**, 761-78.

14. Sandigursky, M. & Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes, *Current biology : CB.* **9**, 531-4.

15. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H. & Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site, *The EMBO journal.* **21**, 3182-91.

16. Wanner, R. M., Castor, D., Guthlein, C., Bottger, E. C., Springer, B. & Jiricny, J. (2009) The uracil DNA glycosylase UdgB of Mycobacterium smegmatis protects the organism from the mutagenic effects of cytosine and adenine deamination, *Journal of bacteriology.* **191**, 6312-9.

17. Srinath, T., Bharti, S. K. & Varshney, U. (2007) Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis, *DNA repair.* **6**, 1517-28.

18. Malshetty, V. S., Jain, R., Srinath, T., Kurthkoti, K. & Varshney, U. (2010) Synergistic effects of UdgB and Ung in mutation prevention and protection against commonly encountered DNA damaging agents in Mycobacterium smegmatis, *Microbiology.* **156**, 940-9.

19. Starkuviene, V. & Fritz, H. J. (2002) A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus, *Nucleic acids research.* **30**, 2097-102.

20. Sakai, T., Tokishita, S., Mochizuki, K., Motomiya, A., Yamagata, H. & Ohta, T. (2008) Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus, *DNA repair.* **7**, 663-9.

21. Lee, H. W., Dominy, B. N. & Cao, W. (2011) New Family of Deamination Repair Enzymes in Uracil-DNA Glycosylase Superfamily, *J Biol Chem.* **286**, 31282-7.

22. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues, *Proceedings of the National Academy of Sciences of the United States of America.* **71**, 3649-53.

23. Huffman, J. L., Sundheim, O. & Tainer, J. A. (2005) DNA base damage recognition and removal: new twists and grooves, *Mutat Res.* **577**, 55-76.

24. Sousa, M. M., Krokan, H. E. & Slupphaug, G. (2007) DNA-uracil and human pathology, *Molecular aspects of medicine.* **28**, 276-306.

25. Stivers, J. T. & Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases, *Chem Rev.* **103**, 2729-59.

26. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase, *Nature.* **373**, 487-93.

27. Cone, R., Bonura, T. & Friedberg, E. C. (1980) Inhibitor of uracil-DNA glycosylase induced by bacteriophage PBS2. Purification and preliminary characterization, *The Journal of biological chemistry.* **255**, 10354-8.

28. Friedberg, E. C., Ganesan, A. K. & Minton, K. (1975) N-Glycosidase activity in extracts of Bacillus subtilis and its inhibition after infection with bacteriophage PBS2, *Journal of virology.* **16**, 315-21.

29. Wang, Z. & Mosbaugh, D. W. (1988) Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in Escherichia coli, *Journal of bacteriology.* **170**, 1082-91.

30. Savva, R. & Pearl, L. H. (1995) Cloning and expression of the uracil-DNA glycosylase inhibitor (UGI) from bacteriophage PBS-1 and crystallization of a uracil-DNA glycosylase-UGI complex, *Proteins.* **22**, 287-9.

31. Takahashi, I. & Marmur, J. (1963) Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for Bacillus subtilis, *Nature.* **197**, 794-5.

32. Wang, Z. & Mosbaugh, D. W. (1989) Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase, *J Biol Chem.* **264**, 1163-71.

33. Bennett, S. E. & Mosbaugh, D. W. (1992) Characterization of the Escherichia coli uracil-DNA glycosylase.inhibitor protein complex, *J Biol Chem.* **267**, 22512-21.

34. Savva, R. & Pearl, L. H. (1995) Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex, *Nature structural biology.* **2**, 752-7.

35. Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA, *Cell.* **82**, 701-8.

36. Beger, R. D., Balasubramanian, S., Bennett, S. E., Mosbaugh, D. W. & Bolton, P. H. (1995) Tertiary structure of uracil-DNA glycosylase inhibitor protein, *J Biol Chem.* **270**, 16840-7.

37. Putnam, C. D., Shroyer, M. J., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W. & Tainer, J. A. (1999) Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase, *Journal of molecular biology.* **287**, 331-46.

38. Karran, P., Cone, R. & Friedberg, E. C. (1981) Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase, *Biochemistry.* **20**, 6092-6.

39. Nakagawa, S., Takai, K., Inagaki, F., Horikoshi, K. & Sako, Y. (2005) Nitratiruptor tergarcus gen. nov., sp. nov. and Nitratifractor salsuginis gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the epsilon-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough, *International journal of systematic and evolutionary microbiology.* **55**, 925-33.

40. Fisher, C. L. & Pei, G. K. (1997) Modification of a PCR-based site-directed mutagenesis method, *BioTechniques.* **23**, 570-1, 574.

41. Xia, B., Liu, Y., Guevara, J., Li, J., Jilich, C., Yang, Y., Wang, L., Dominy, B. N. & Cao, W. (2017) Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily, *Scientific reports.* **7**, 45978.

42. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes, *Acta Crystallogr D Biol Crystallogr.* **62**, 859-66.

43. Pannu, N. S., Waterreus, W. J., Skubak, P., Sikharulidze, I., Abrahams, J. P. & de Graaff, R. A. (2011) Recent advances in the CRANK software suite for experimental phasing, *Acta Crystallogr D Biol Crystallogr.* **67**, 331-7.

44. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D Biol Crystallogr.* **68**, 352-67.

45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr.* **66**, 486-501.

46. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr.* **66**, 12-21.

47. King, K., Benkovic, S. J. & Modrich, P. (1989) Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease, *The Journal of biological chemistry.* **264**, 11807-15.

48. Vermote, C. L. & Halford, S. E. (1992) EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition, *Biochemistry.* **31**, 6082-9.

49. Guex, N. & Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis.* **18**, 2714-23.

50. Humphrey, W., Dalke, A. & Schulten, K. (1996) VMD: Visual molecular dynamics, *Journal of Molecular Graphics.* **14**, 33-38.

51. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *Journal of Physical Chemistry B.* **102**, 3586-3616.

52. MacKerell, A. D. & Banavali, N. K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution, *Journal of Computational Chemistry.* **21**, 105-120.

53. Darden, T., York, D. & Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, *Journal of Chemical Physics.* **98**, 10089-10092.

54. Adelman, S. A. & Doll, J. D. (1976) Generalized Langevin equation approach for atom/solid‐surface scattering: General formulation for classical scattering off harmonic solids, *Journal of Chemical Physics.* **64**, 2375-2388.

55. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005) Scalable molecular dynamics with NAMD, *Journal of Computational Chemistry.* **26**, 1781-1802.

56. Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic acids research.* **42**, W320-4.

57. Pang, P., Yang, Y., Li, J., Wang, Z., Cao, W. & Xie, W. (2017) SMUG2 DNA glycosylase from Pedobacter heparinus as a new subfamily of the UDG superfamily, *The Biochemical journal.* **474**, 923-938.

58. Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A. & Krokan, H. E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase, *EMBO J.* **15**, 3442-7.

59. Chembazhi, U. V., Patil, V. V., Sah, S., Reeve, W., Tiwari, R. P., Woo, E. & Varshney, U. (2017) Uracil DNA glycosylase (UDG) activities in Bradyrhizobium diazoefficiens: characterization of a new class of UDG with broad substrate specificity, *Nucleic acids research.* **45**, 5863-5876.

60. Nayek, A., Sen Gupta, P. S., Banerjee, S., Mondal, B. & Bandyopadhyay, A. K. (2014) Salt-bridge energetics in halophilic proteins, *PloS one.* **9**, e93862.

61. Petukh, M., Li, M. & Alexov, E. (2015) Predicting Binding Free Energy Change Caused by Point Mutations with Knowledge-Modified MM/PBSA Method, *PLoS Comput Biol.* **11**, e1004276.

Note:

This chapter has been published as a peer-reviewed paper:

Jing Li*, Ran Chen*, Ye Yang*, Zhemin Zhang, Guang-Chen Fang, Wei Xie, Weiguo Cao. An Unconventional Family 1 Uracil DNA Glycosylase in *Nitratifractor salsuginis*. *FEBS J.* 2017 doi: 10.1111/febs.14285. *, These authors contribute equally

CHAPTER FOUR

A URACIL DNA GLYCOSYLASE FROM *JANTHINOBACTERIUM AGARICIDAMNOSUM* AS A POTENTIAL EVOLUTIONARY INTERMEDIATE

I. Abstract

Uracil DNA glycosylase (UDG) plays a critical role in DNA repair by cleavage of uracil from genomic DNA. So far, all the 6 UDG families share a similar α/β fold structure. Especially, family 1 UNG and family 4 UDGa have a similar topological structure and catalytic center, and both of them are exclusive UDG with no activity on other deaminated bases. In the present study, we characterized another exclusive UDG enzyme from *Janthinobacterium agaricidamnosum* (Jag UNG). Phylogenetic study showed that, like the recently reported Nsa UNG, Jag UNG belongs to a subgroup of family 1 UNG. Similarly with Nsa UNG, Jag UNG harbors a salt bridge network with uracil-containing DNA backbone. A single mutant A82E increases the enzyme efficiency by 200-fold by stabilizing the salt bridge. In addition, E76 and G77 in motif 1 of Jag UNG are functional correlated. Replacement the doublet E76G77 with Q76D77 increased the enzyme efficiency by 30-fold while any single mutant (E76Q or G77D) reduce or inactivate the enzyme. In motif 2, L245 locates near the interface between protein and DNA, a substitution with the positively charged His increase the enzyme efficiency by changing the electrostatic potential of the binding surface. Meanwhile, Jag UNG can't be inhibited by the UDG inhibitor (Ugi) as the intruding loop in motif 2 is too short. By this work, we have a better understanding about the catalytic and evolutionary mechanisms of UDG enzymes.

II. Introduction

Uracil frequently appears in DNA either by cytosine deamination or uracil mis-incorporation. The presence of uracil may result in the C to T transition mutation as uracil tends to pair with adenine. Uracil DNA glycosylase (UDG) is one of the essential DNA repair enzymes that recognize and excise uracil from DNA. To date, 6 different UDG families have been comprehensively investigated. Each family exhibits different substrate specificity and enzyme efficiency. Overall, all the enzymes in the UDG superfamily are active to uracil except the enzymes in family 6 which act as hypoxanthine DNA glycosylase (HDG) [1-12].

Structurally, all of UDG enzymes shared a similar α/β fold sandwich architecture (A few β-sheets located in the center and several α-helixes around) [10-15]. Evolutionarily speaking, all UDG enzymes come from a common ancestor. Especially, family 1 UNG and family 4 UDGa have a similar overall topological structure and catalytic center, therefore, both of them are robust but exclusive UDG. However, the enzyme efficiency of family 1 UNG and family 4 UDGa are significantly different as the slight difference in their catalytic site. Another distinct difference between them is related to the effect of uracil glycosylase inhibitor (Ugi) that is able to specifically and tightly bind to the DNA binding site of family 1 UNG [16, 17]. However, Ugi has no effect to family 4 UDGa. Given such high similarity but distinct features of family 1 UNG and family 4 UDGa, they are believed to be evolutionary related and it is interesting to study the evolutionary process between them.

Previously, based on the phylogenetic analysis and phyletic distribution of UDG superfamily, family 4 UDGa was proposed as the common ancestor of other UDG families and family 1 UNG is evolved from family 4 UDGa by acquiring different catalytic residues [18]. Recently, we characterized an untraditional family 1 UNG from *Nitratifractor salsuginis* (Nsa UNG) (Chapter 3). Representing a new group of family 1 UNG, Nsa UNG acquired the catalytic residues of family 1 UNG, but did not evolve a proper interface for Ugi binding. Only distributed in the genus of *Nitratifractor* and *Sulfurovum,* Nsa UNG-like enzymes probably are evolutionary intermediates from family 4 UDGa and family 1 UNG.

In present study, through a large scale of database searching, we found another group of evolutionary intermediates between family 1 UNG and family 4 UDGa from *Oxalobacteraceae* family, represented by the enzyme from *Janthinobacterium agaricidamnosum* (Jag UNG). Through phylogenetic, enzymatic and structural studies of Jag UNG, and comparison with the Nsa UNG, family 1 UNG and family 4 UDGa, we partly explain the evolutionary mechanisms from family 4 UDGa to family 1 UNG.

III. Materials and Methods

Reagents, media, and strains

All routine chemical reagents were purchased from Fisher Scientific (Suwanee, GA) or VWR (Suwanee, GA). Restriction enzymes, Phusion DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). dNTPs were purchased from Promega (Madison, WI). Gel DNA recovery Kit was purchased from Generay (Shanghai, China). Oligodeoxyribonucleotides were ordered from Eurofins

Genomics (Huntsville, AL). The LB medium was prepared according to standard recipes. Hi-Di™ Formamide and GeneScan 500 LI Size Standard were purchased from Applied Biosystems®.

Cloning, site-directed mutagenesis, Expression and Purification of Jag UNG

The Jag UNG gene from *Janthinobacterium agaricidamnosum* (GenBank accession number: CDG83998.1) was amplified by PCR using the forward primer Jag UNG F (5'- GGGAATTCC ATATGTCGACTTCCCCGCAT-3'; the NdeI site is underlined) and the reverse primer Jag UNG R (5'-CCCAAGCTTAACCTGCAACAAATGCAT-3'; the HindIII site is underlined).The PCR reaction mixture (50 µl) consisted of 20 ng *Janthinobacterium agaricidamnosum* genomic DNA, 500 nM forward and reverse primers, 1 x Phusion DNA polymerase buffer, 200 µM each dNTP and 0.2 unit of Phusion DNA polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 98°C for 30 s; 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 s, annealing at 65°C for 20 s, and extension at 72°C for 30 s; and a final extension step at 72°C for 3 min. The PCR product was purified by gel DNA recovery kit (Generay). The purified PCR product and plasmid pET21a were digested by NdeI and HindIII, purified by gel DNA recovery kit, and ligated according to the manufacturer's instructional manual. The ligation mixture was transformed into *E. coli* strain HB101 competent cells by electroporation. The sequence of the Jag UNG gene in the resulting plasmid (pET21a-Jag UNG) was confirmed by DNA sequencing. The resulting plasmid with wild-type Jag UNG was used as the template plasmid for all other Jag UNG mutants. Amplification of

the mutant DNA and DpnI mediated site-directed mutagenesis procedures were modified as previously described by using primers carrying the desired mutations [19]. Briefly, PCR mixtures (25 µl) contained 10 ng of pET-21a (+)-Jag UNG as a template, 65 nM of each primer pair, 200 µM each dNTP, 1 x Phusion PCR polymerase buffer, and 1 unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation step at 98°C for 2 min; 25 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for 5 min; and a final extension step at 68°C for 10 min. After treatment with 2 units of DpnI for 1 h at 37°C, 5-µl PCR products were transformed into *E. coli* DH5α competent cells. Successful insertion and mutation in the resultant clones were confirmed by DNA sequencing. The pET21a-Jag UNG WT and mutants were transformed into *E. coli* strain BH214 (*ung*$^-$, *mug*$^-$) by the standard protocol to express the C-terminal His-6-tagged Jag UNG protein. Induction, sonication and purification were carried out as previously described [10].

Cloning, Site-directed mutagenesis, Expression and Purification of Nsa UNG

Cloning, site-directed mutagenesis, Expression and Purification of Nsa UNG were carried out as previously described in Chapter 3.

Oligodeoxynucleotide Substrates

Oligodeoxynucleotides containing deoxyuridine(U), deoxyinosine (I) and deoxyxanthosine (X) were obtained or constructed as previously described [7].

DNA Glycosylase Activity Assay

DNA glycosylase cleavage assays for JagUDG were performed at optimized temperature 37°C for 60 min in a 10 µl reaction mixture containing 10 nM

oligonucleotide substrate, 100 nM Jag UNG, 20 mM Tris-HCl buffer (pH 7.5), 1mM DTT, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95°C for 5 min after adding 1 µl of 1N NaOH. The reaction mixtures (2 µl) were mixed with 7.8 µl Hi-D formamide and 0.2 µl GeneScan 500 LI Size Standard and analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

Enzyme kinetic analysis

Uracil DNA glycosylase assays were performed at optimized condition with 10 nM G/U substrates with enzyme in excess ranging from 50 nM to 2000 nM. Samples were collected at 1 min, 2.5 min, 5 min, 10 min, 15 min, 25 min, 30 min, 40 min and 60 min. The apparent rate constants for each concentration were determined by curve fitting using the integrated first-order rate equation (1):

$$P = P_{max}(1 - e^{-k_{obs}t}) \quad (1)$$

Where P is the product yield, P_{max} is the maximal yield, t is time and k_{obs} is the apparent rate constant.

The kinetic parameters k_2 and K_m were obtained from plots of k_{obs} against the total enzyme concentration ($[E_0]$) using a standard hyperbolic kinetic expression with the program GraphPad 4.1 following the equation (2)[20]

$$k_{obs} = \frac{k_2\ [E_0]}{K_m +\ [E_0]} \quad (2)$$

For some mutants with a large K_m in which $K_m >> [E_0]$, the kinetic parameter k_2/K_m values were obtained from plots of k_{obs} against total enzyme concentration ($[E_0]$) using a linear regression with program GraphPad 4.1 following the equation (3)[21].

$$k_{obs} = \frac{k_2\ [E_0]}{K_m} \quad (3)$$

Molecular modeling

The structure of Jag UNG was predicted by SWISS-Model with Nsa UNG structure (PDB 5X3G) as a template [22]. The final predicted structure was refined by Princeton_TIGRESS (server) [23]. The structures of Jag UNG mutants (A82E, E76Q, G77D, L244H, E76Q/G77D/L244H) were constructed and minimized by UCSF Chimera 1.10.2 [24].

Molecular dynamics simulations

After building the initial complex structures, an explicit solvent system using the TIP3P water model was constructed in the VMD 1.9.2 using a suitably sized box. The minimum distance between any of the atoms of the solvated Jag UNG-DNA complex and the box boundary was maintained to at least 9 Å. Potassium chloride ions were added to the system to achieve an electrically neutral system. The CHARMM 27 all hydrogen force field for proteins 3 and nucleic acids 4 were used. Particle-mesh Ewald summation 5 was applied in the periodic boundaries condition for the efficient calculation of long-range electrostatic interactions. Energy minimization was performed by using 8000 steps to remove any unfavorable van der Waals clashes while minimally perturbing the original modeled structure. Using a Langevin barostat 6, an isothermal-isobaric ensemble (NPT) was constructed in NAMD program 7 and the system was heated gradually from 100 K to 310 K over a period of 400 ps. An integration time step of 2 fs was used in order to avoid any significant structural deformation during heating, equilibration and production runs. Calculated RMSD demonstrated that the trajectories have stabilized in 5 ns for each

simulation (data not shown). Thus a total of 5 ns equilibration followed by 100 ns production simulation were performed for each structural analysis. VMD 1.9.2 was used for visualization and further analysis. RMSF calculations were performed for Cα atoms of Jag UNG. A hydrogen bond was considered to be formed when the distance between the donor and acceptor is less than 3.0 Å and also the angle between the donor hydrogen and acceptor is less than 20 degrees.

IV. Results

Phylogenetic analysis of Jag UNG

A BLAST analysis showed that Jag UNG has high degree of similarity to family 1 UNG. However, according to the multiple sequence alignment of the functional motifs of Jag UNG and other UDG families, Jag UNG looks like a hybrid enzyme of family 1 UNG and family 4 UDGa. In one way, Jag UNG contains the family 4 unique Glu and Gly (E76 and G77) doublet in motif 1. In another way, Jag UNG also contains the family 1 unique Tyr (Y79) in motif 1 (Fig. 4.3A). Whole sequence alignment showed that Jag UNG shares 30.03%, 55.72% and 28.17% similarity (23.14%, 40.62% and 19.51% identity) with family 1 Eco UNG, Nsa UNG and family 4 Tth UDGa, respectively (Fig. 4.1). To illuminate the evolutionary status of Jag UNG, we did a phylogenetic analysis of UDG superfamily including Jag UNG and its homologs. The result showed that Jag UNG and its homologs have formed a distinct clade from conventional family 1 UNG (Fig.4.2). Together with Nsa UNG-like enzymes, they are another branch of family 1 UNG.

```
JagUNG       MSTSPHS---L-----IPASILEA------LALADVSWRPILQQ---------GLQAVAS
NsaUNG       MTNNKYTNRPI-----HQLQELLR------LNGVDEEWEPILLP---------ALMTLED
EcoUNG       -------------------------------MANELTWHDVLAEEKQQPYFLNTLQTVAS
HsaUNG       PPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVAE
TthUDGa      ------------------------------------------------MTLELLQAQAQ
                                                                :    .

JagUNG       ADPAYLPALAIDDYLPTEQRLFAAFA-LPIDAVRYVLVGEGPYPRAASATGVCFMDGA--
NsaUNG       SYLEWM--AAGEGYIPPRDRLLAAFSTLRPNEVRYILFGQDPYPRPESAIGYAFIDGR--
EcoUNG       ER------QSGVTIYPPQKDVFNAFRFTELGDVKVVILGQDPYHGPGQAHGLAFSVRP--
HsaUNG       ER-------KHYTVYPPPHQVFTWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQR--
TthUDGa      NCT-------ACRLMEGRTR--VVFG-EGNPDAKLMIVGEGPGEEED-KTGRPFVGKAGQ
                                       .: ::.*:.*        *  *

JagUNG       -VGSLWSEAGLSKQVNRATSLRNFMKMLLVADGQLTAGDTSGAAMAPVSARARGNGSGTI
NsaUNG       -VREIFSPRGLSREVNRATSLRNFIKMALVARGSLDPRDTSQEAIAALDK------TLLV
EcoUNG       -------------GIAIPPSLLNMYKELENTI------------------------PGFT
HsaUNG       -------------PVPPPPSLENIYKELSTDI------------------------EDFV
TthUDGa      LLNRILEAAGIPREEVY---ITNIVKCRPPQN------------RAPLPD----------
                           : *: *

JagUNG       QTLPDLQHNLSHQGFLLLNAALVFR----------AHVAPVHD-ARGWLPFLQVVL-A--
NsaUNG       SQMRELRENFERSGVLLLNMALLFT----------SKEESRRH-IRAWRAFIEKLL-E--
EcoUNG       RPNHGYLESWARQGVLLLNTVLTVR----------AGQAHSHA-SLGWETFTDKVI-S--
HsaUNG       HPGHGDLSGWAKQGVLLLNAVLTVR----------AHQANSHK-ERGWEQFTDAVV-S--
TthUDGa      -EAKICTDKWLLKQIELIAPQIIVPLGAVAAEFFLGEKVSITKVRGKWYEWHGIKVFPMF
                        . . *:   : .           .           *  :    :

JagUNG       --ALAERSESAGRTAPTLVLWGKIAELL---KKLPVTARLPQVAAEHPYNLTFINNTTMQ
NsaUNG       --GF-------EAYGPTLILFGAHAREV---QKLKSARGLPQVALEHPYNHTFIVNEKAW
EcoUNG       --LINQ-----HREGVVFLLWGSHAQKK---GAIIDKQRHHVLKAPHPSPLSAH-----R
HsaUNG       --WLNQ-----NSNGLVFLLWGSYAQKK---GSAIDRKRHHVLQTAHPSPLSVY-----R
TthUDGa      ---------------------------------------------HPAYLLRNPSH---
                                                           **

JagUNG       ELFGPMHLLQV-----------------------
NsaUNG       ELFGPMDLLLKR----------------------
EcoUNG       GFFGCNHFVLANQWLEQRGETPIDWMPVLPAESE
HsaUNG       GFFGCRHFSKTNELLQKSGKKPIDWKEL------
TthUDGa      RAPGSPKHLTWLDIQEVKRALDALPPKERRPVKAVSQEPLF
```

Figure 4. 1 Sequence alignment of Jag UNG and other family 1 UNG and family 4 UDGa. Jag UNG: Jag, *Janthinobacterium agaricidamnosum*, CDG83998.1; Nsa UNG: Nsa, *Nitratifractor. salsuginis*, YP_004167197.1; Eco, *E. coli*, NP_289138; Hsa, *H. sapiens*, NP_003353; Tth, *T. thermophilus* HB27, YP_004341.1.

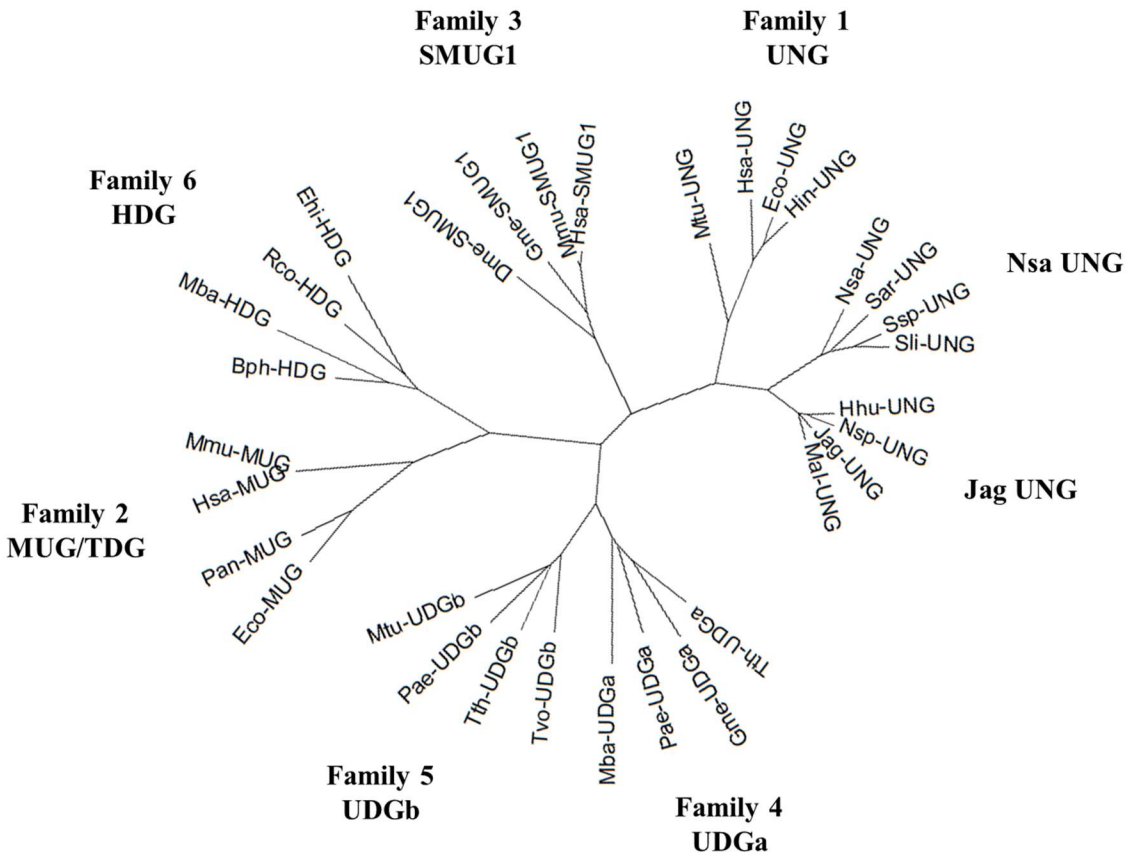


Figure 4. 2 Phylogenetic Analysis of Uracil DNA glycosylase superfamily. The phylogenetic analysis was performed using the neighbor-joining method in MEGA 6. GenBank accession numbers are shown after the species names. Family 1 (UDG): Eco, *E. coli*, NP_289138; Hin, *Haemophilus influenzae* KR494, YP_008544610.1; Mtu, *Mycobacterium tuberculosis*, WP_003908950.1; Hsa, *H. sapiens*, NP_003353. Family 2 (MUG/TDG): Eco, *E. coli*, P0A9H1; Pan, *Pantoea ananatis* LMG 20103, ADD78558.1; Hsa, *H. sapiens*, NP_003202; Mmu, *Mus musculus*, XP_003945901.1. Family 3 (SMUG1): Gme, *G. metallireducens* GS-15,YP_383069; Hsa, *Homo sapiens*, NP_055126; Mmu, *Mus musculus*, NP_082161; Dme, *Drosophila melanogaster*, NP_650609.1. Family 4 (UDGa): Tth, *T. thermophilus* HB27, YP_004341.1. ; Pae, *P. aerophilum str.* IM2, NP_558739.1; Gme, *Geobacter metallireducens* GS-15, YP_006721625.1; Mba, *Methanosarcina barkeri str*. Fusaro, YP_305330.1. Family 5 (UDGb): Tth, *T. thermophilus* HB8, YP_144415.1; Pae, *P. aerophilum str*. IM2, NP_559226; Tvo, *Thermoplasma volcanium* GSS1, NP_111346.1; Mtu, *M. tuberculosis* H37Rv, P64785 (Rv1259). Family 6 (HDG): Bph, Burkholderia phymatum STM815, YP_001858334.1; Mba, *Methanosarcina barkeri str*. Fusaro,YP_304295.1; Rco, *Ricinus communis,* XP_002536323.1; Ehi, *Entamoeba histolytica* HM-1:IMSS, XP_655177.1. Jag UNG: Jag, *Janthinobacterium agaricidamnosum*, CDG83998.1; Hhu,

Herbaspirillum huttiense, WP_039784854.1; Nsp, *Noviherbaspirillum sp*. Root189, WP_057292635.1; Mal, *Massilia alkalitolerans*, WP_027867664.1; Nsa UNG: Nsa, *Nitratifractor. salsuginis*, YP_004167197.1; Ssp, *Sulfurovum sp*. AR, WP_008244665.1; Sar, *Sulfurospirillum arcachonense*, WP_024954916.1; Sli, *Sulfurovum lithotrophicum*, WP_046550169.1.

Enzyme properties of Jag UNG

To investigate the DNA repair functions of Jag UNG, we examined the enzyme properties of JagUDG *in vitro* by testing their DNA glycosylase activity on uracil- (U), hypoxanthine- (Inosine, I) and xanthine- (X) containing DNA substrates (Fig. 4.3B and 4.3C). The assay results indicated that Jag UNG can't excise hypoxanthine or xanthine from DNA, instead, it only can recognize and cleave uracil from double-stranded DNA, in the order of G/U> A/U > C/U > T/U (Fig. 4.3D). Therefore, we concluded that just like family 1 UNG, Nsa UNG and family 4 UDGa, Jag UNG is an exclusive UDG. However, given the lower enzyme efficiency (k_2/K_m is only 3×10^{-4} min^{-1} nM^{-1}), Jag UNG works much slower than family 1 UNG (human UNG 34 min^{-1} nM^{-1}), Nsa UNG (48 min^{-1} nM^{-1}) and family 4 UDGa (Tth UDGa 1.4×10^{-2} min^{-1} nM^{-1}) (Table 4.1) [25].

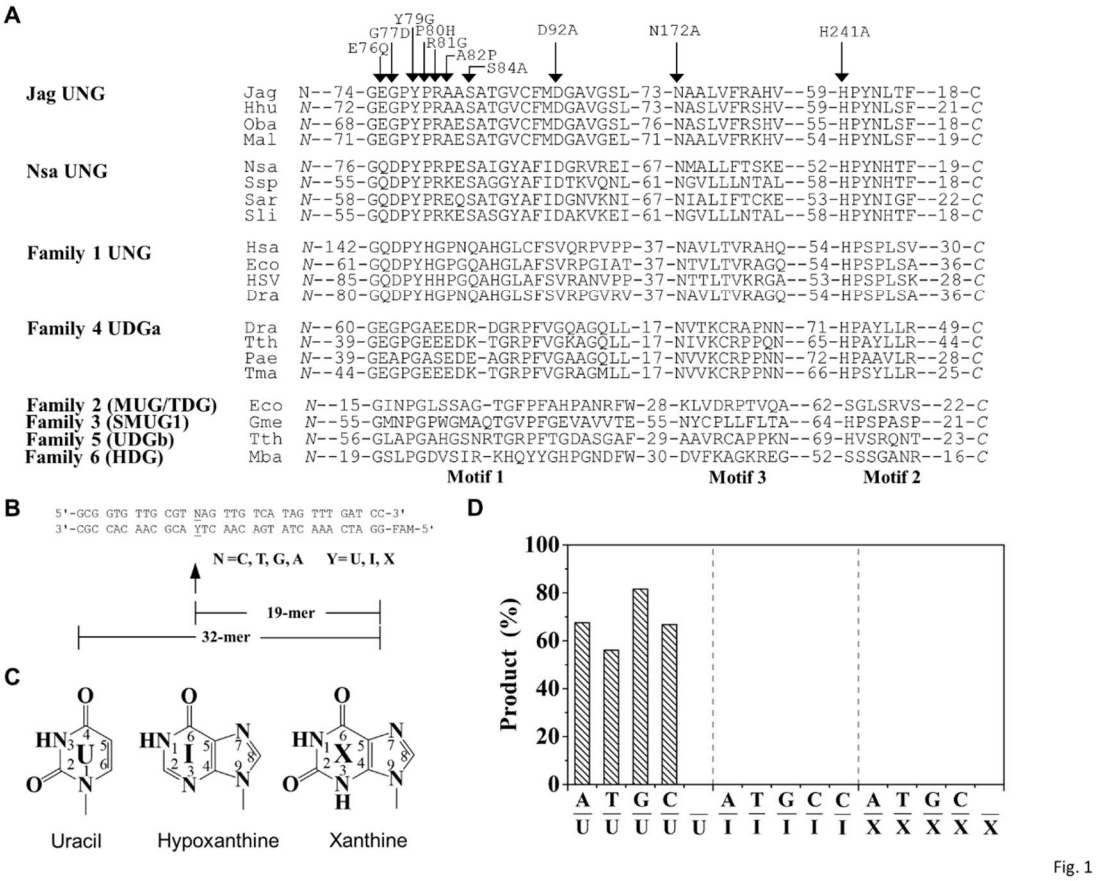


Fig. 1

Figure 4. 3 Sequence alignment, substrate and UDG activity of Jag UNG. A. Sequence alignment of Jag UNG and other UDG families; Jag UNG: Jag, *Janthinobacterium agaricidamnosum*, CDG83998.1; Hhu, *Herbaspirillum huttiense*, WP_039784854.1; Oba, *Oxalobacteraceae bacterium* AB_14, WP_026334292.1; Mal, *Massilia alkalitolerans*, WP_027867664.1; Nsa UNG: Nsa, *Nitratifractor. salsuginis*, YP_004167197.1; Ssp, *Sulfurovum sp*. AR, WP_008244665.1; Sar, *Sulfurospirillum arcachonense*, WP_024954916.1; Sli, *Sulfurovum lithotrophicum*, WP_046550169.1; Family 1 (UNG): Hsa, *H. sapiens*, NP_003353; Eco, *E. coli*, NP_289138; HSV, *Herpes Simplex Virus 1*, P10186.1. Dra, *Deinococcus radiodurans* R1, NP_294412; Family 4 (UDGa): Dra (DR 1751), *D. radiodurans* R1, NP_295474; Tth, *Thermus thermophilus* HB27, YP_004341.1; Pae, *P. aerophilum str*. IM2, NP_558739.1; Tma, *Thermotoga maritima* MSB8, NP_228321.1; Family 2 (MUG/TDG): Eco, *E. coli*, P0A9H1; Family 3 (SMUG1): Gme, *G. metallireducens* GS-15, YP_383069; Family 5 (UDGb): Tth, *T. thermophilus* HB8, YP_144415.1; Family 6 (HDG): Mba, *Methanosarcina* barkeri str. Fusaro, YP_304295; B. Sequence of uracil, hypoxanthine, xanthine-containing DNA substrates; C. Chemical structures of uracil, hypoxanthine and xanthine; D. DNA glycosylase activity of WT Jag UNG on U, I, X-containing DNA substrates. The assay

was performed as described in Experimental Procedures under DNA glycosylase activity assay.

Table 4. 1 Kinetic constants of Jag UNG, NsaUNG, human UNG and TthUDGa on G/U substrate[a]

Enzyme	K_m (nM)	k_2 (min^{-1})	k_2/K_m (min^{-1} nM^{-1})
Jag UNG-WT	N.D.[b]	N.D.	3×10^{-4}
Jag UNG-E76Q	N.A.[c]	N.A.	N.A.
Jag UNG-G77D	N.D.	N.D.	2×10^{-5}
Jag UNG-E76Q/G77D	528±100	4.78±0.32	9×10^{-3}
Jag UNG-E76Q/G77D/L245H	197±53	12±0.8	6×10^{-2}
Jag UNG-P80H	N.A.	N.A.	N.A.
Jag UNG-R81G	N.A.	N.A.	N.A.
Jag UNG-A82E	428±115	30±2.5	6×10^{-2}
Jag UNG-D92A	N.A.	N.A.	N.A.
Jag UNG-S84A	N.D.	N.D.	5×10^{-5}
Jag UNG-H241A	N.D.	N.D.	N. D.
Jag UNG-L245H	N.D.	N.D.	1.8×10^{-3}
Nsa UNG-WT	27 ± 2.9	1300 ± 120	48
Nsa UNG-H234L	675 ± 119	0.44 ± 0.03	6.52×10^{-4}
human UNG-WT	169 ± 25	5820 ± 330	34
Tth UDGa-WT	970 ± 240	13.8 ± 0.6	1.4×10^{-2}

[a]: The reactions were performed as described in Experimental Procedures under enzyme kinetic analysis. Data are an average of three independent experiments.
[b]: Not determined. Individual K_m and k_2 values were not determined due to a relative large K_m.
[c]: No activity.

Primary sequence alignment showed that Jag UNG contains the same catalytic residues in the functional motifs (E76 and G77 in motif 1, H241 in motif 2 (Fig. 4.2A)) as

family 4 UDGa (E41, G42 and H155 in TthUDGa). Structurally, those residues are located with similar orientations in the catalytic center and may play a similar role in catalysis (Fig. 4.4). Consistent with the studies in family 4 UDGa, mutagenesis of any of these residues in Jag UNG results in the loss of enzyme activity (Table 4.1). Besides the similarity with family 4 UDGa, Jag UNG also has a similar salt bridge network with DNA as Nsa UNG. Arg 81 directly interacts with the phosphate 5' to the deoxyuridine and the side chain of Arg 81 formed a salt bridge with the invariant Asp 92 and the side chain of Asp 92 was stabilized by hydrogen bonding with the side chain of Ser 84. P80 has an important role in stabilizing the salt bridge. (Fig. 4.5). Any mutants (P80H, A81G, D92A, and S84A) which destroyed the salt bridge network, either completely inactivate the enzyme or significantly reduced the enzyme efficiency (Table 4.1).

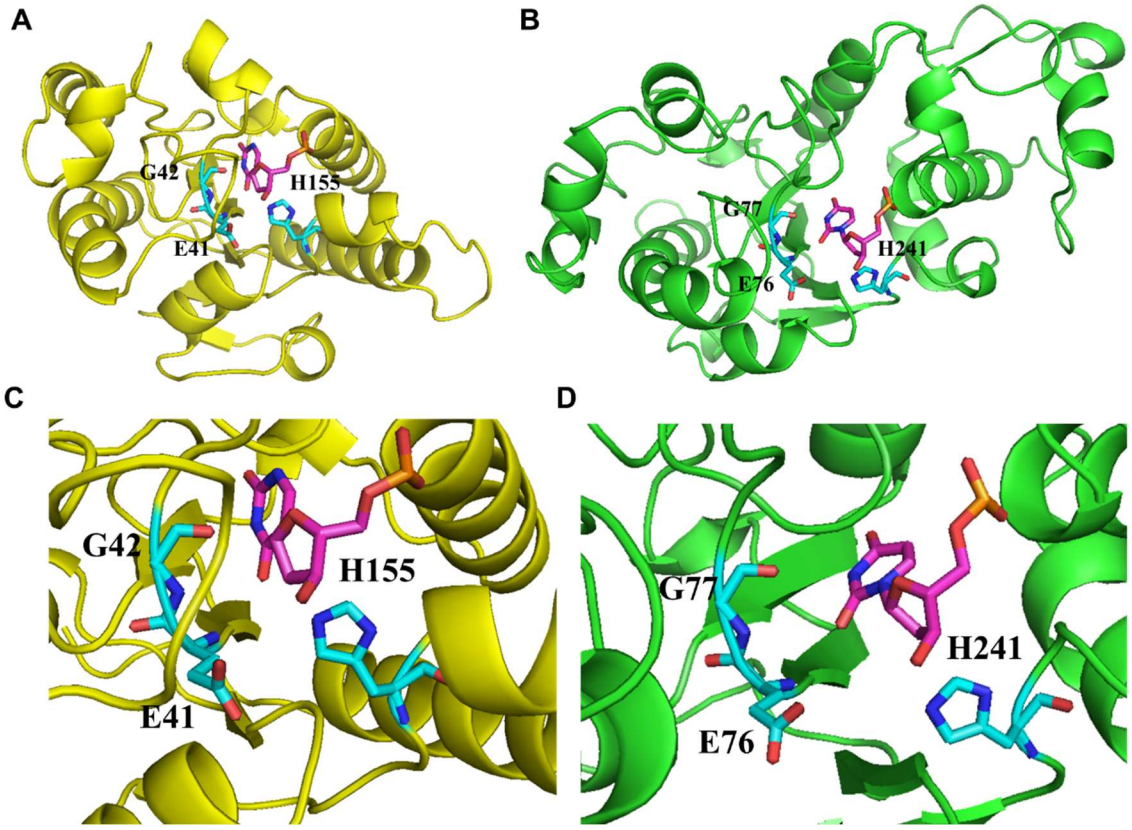


Figure 4. 4 Comparison the catalytic center between Family 4 UDGa (TthUDGa) and Jag UNG. A. Active site of Tth UDGa (PDB 1UI0). Tth UDGa structure is shown in cartoon mode and catalytic residues (E41, G42 and H155) are shown in licorice. B. Active site of Jag UNG. Jag UNG structure is shown in cartoon mode and catalytic residues (E76, G77 and H241) are shown in licorice. C. Close-up view of active site of Tth UDGa (PDB 1UI0). Tth UDGa structure is shown in cartoon mode and catalytic residues (E41, G42 and H155) are shown in licorice. D. Close-up view of active site of Jag UNG. Jag UNG structure is shown in cartoon mode and catalytic residues (E76, G77 and H241) are shown in licorice.

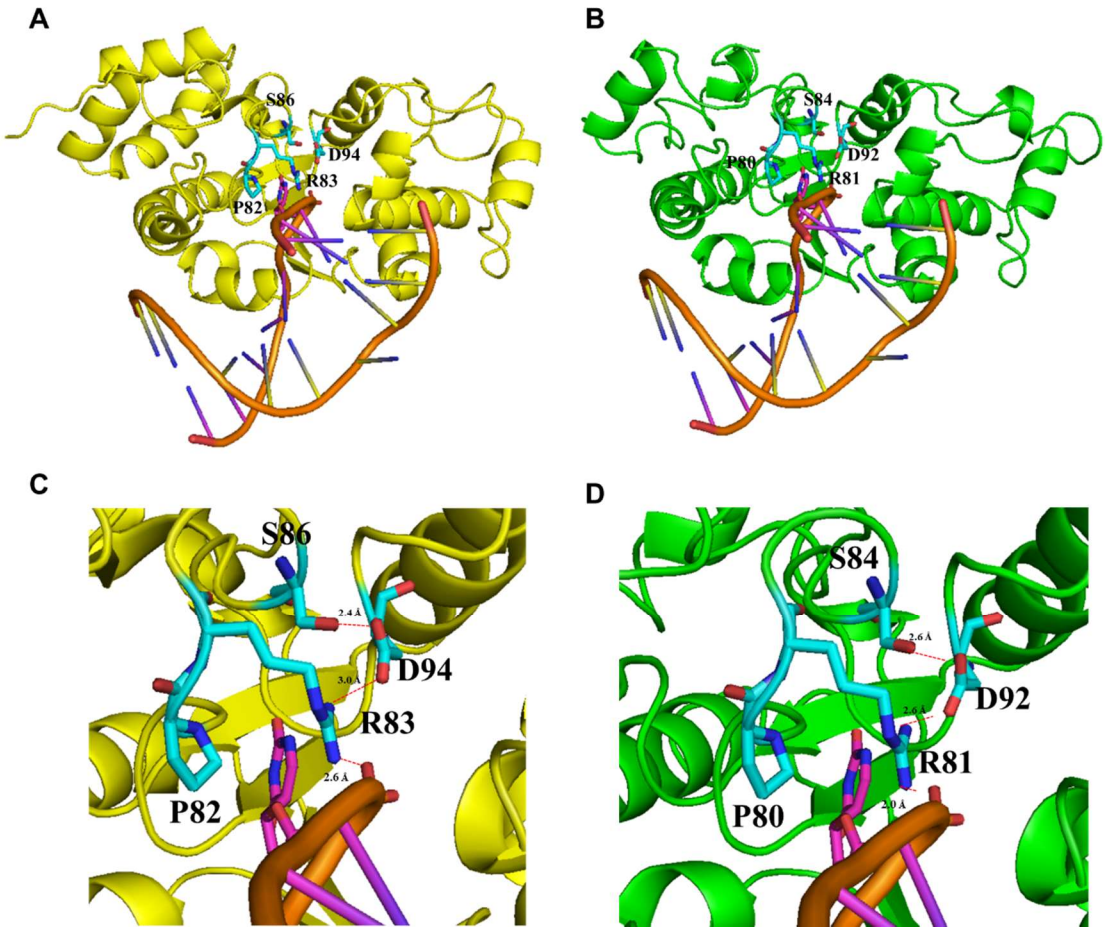


Figure 4. 5 Comparison of the salt bridge network in Nsa UNG and Jag UNG. A. The interactions between Nsa UNG-R83 and modeled uracil-containing DNA (DNA model was taken from PDB 1EMH). B. The interactions between Jag UNG and Uracil-containing DNA (PDB 1EMH). C. Close-up view of salt bridge network in Nsa UNG that involves in DNA phosphate backbone interaction. The network of side chain interactions among P82, R83, S86 and D94 are shown in licorice; D. Close-up view of the salt bridge network in Jag UNG that involves in DNA phosphate backbone interaction. The network of side chain interactions among P80, R81, S84 and D92 are shown in licorice.

Replacement of EG with QD

Previously, site-directed mutagenesis work, enzyme kinetics, molecular dynamics (MD) simulation and mutual information analysis proved the residues E and G in the motif 1 of family 4 UDGa as well as the residues Q and D in the motif 1 of family 1

UNG are functionally correlated, respectively [25]. In present study, a replacement of E76G77 with Q76D77 in Jag UNG results in a 30-fold increase of the enzyme efficiency and significantly lower K_m. Either the single mutation E76Q or G77D lost or reduce most UDG activity (Table 4.1). This result is consistent with the previous study and further confirmed their correlation [25]. MD simulation analysis demonstrated that the average number of hydrogen bond between the double mutant protein (Jag UNG-E76Q/G77D) and uracil-containing DNA increased to 5.6 which is significantly higher than that of WT (4.8), Jag UNG-E76Q (4.3) and Jag UNG-G77D (4.6) proteins (Table 4.2). This stronger hydrogen bond may explain the elevated enzyme efficiency and substrate binding affinity of Jag UNG-E76Q/G77D.

Table 4. 2 Average number of Hydrogen Bond of wild type and mutant Jag UNG proteins[a].

Jag UNG	Avg Number of Hydrogen Bond between Jag UNG-DNA
WT	4.8 ± 0.1
A82E	5.3 ± 0.1
L245H	5.2 ± 0.1
E76Q	4.3 ± 0.1
G77D	4.6 ± 0.1
E76Q/G77D	5.6 ± 0.1
E76Q/G77D/L245H	6.2 ± 0.1

[a]: Avg Number of hydrogen bond between Jag UNG-DNA was calculated using the Hbonds Plugin in VMD.

A82E

In family 4 UDGa enzymes, a highly conserved Glu (E47 in Tth UDGa), located in a short helix, is considered to play a similar role as the Tyr (Y in *E. coli* UNG) in

family 1 UNG. Both of these two residues block the C5 position of uracil base and make the enzyme binding pocket exclusive to uracil. Therefore, family 1 UNG and family 4 UDGa are exclusive uracil DNA glycosylase and have no activity to thymine and other purine substrates [11, 26]. As a potential intermediate between family 4 UDGa and family 1 UNG, Jag UNG has lost the Glu and evolved a Tyr 79 in motif 1. Another conserved Ala from a loop, replaces the Glu at the position 82 in motif 1 (Fig. 4.3A). Interestingly, if we mutated A82 to E and mimic Jag UNG as family 4 UDGa, the K_m is significantly decreased and the enzyme efficiency of Jag UNG-A82E is 200-fold greater than that of Jag UNG-WT (Table 4.1). This result indicates that E82 has a crucial role in the substrate recognition and catalysis in Jag UNG.

L245H

In motif 2 of Jag UNG, a conserved Leu (L245) locates in an α-helx. However, Nsa UNG contains a conserved His (H234) in the corresponding site (Fig. 4.3A). Interestingly, a substitution of L245 with H in Jag UNG increased the enzyme efficiency by 5-fold. Consistently, the enzyme efficiency of the triple mutant Jag UNG-E76Q/G77D/L245H is 7-fold greater than that of double mutant Jag UNG-E76Q/G77D. In addition, the replacement of H234 in Nsa UNG with L decreases the enzyme efficiency thousands-fold and increase the K_m by 25 times (Table 4.2). All those results indicate that His plays a very important role in the catalysis and substrate recognition in both Nsa UNG and Jag UNG.

Ugi effects on Jag UNG

A uracil DNA glycosylase assay in presence of Ugi indicated that, just like Nsa UNG, Jag UNG can't be inhibited by the uracil DNA glycosylase inhibitor (Ugi) (Fig. 4.6A). A native page gel binding assay showed that Ugi can't bind to Jag UNG (Fig. 4.6B). This suggested that Jag UNG has not evolved the specific interface for Ugi binding.

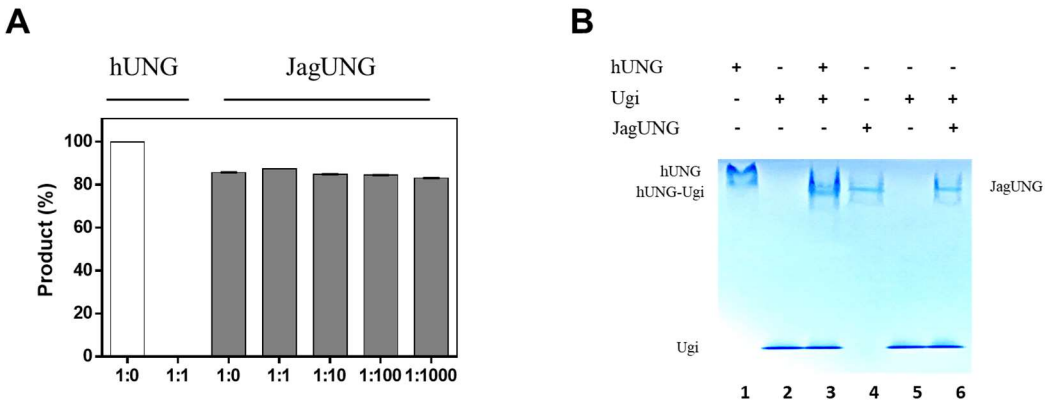


Figure 4. 6 Ugi inhibition and binding analysis. A. The reactions were performed as described under DNA golycosylase assay in Material and Methods with addition of indicated amount of Ugi protein. The concentrations of Jag UNG versus Ugi are shown in molar ratio. B. Image of 15% native PAGE gel analysis showing the binding of Ugi to hUNG and Jag UNG. Lane 1: human UNG (0.1 nmol); Lane 2: Ugi (1 nmol); lane 3: human UNG: Ugi = 1:10; Lane 4: Jag UNG (0.1 nmol); Lane 5: Ugi (1 nmol); lane 6: Jag UNG: Ugi = 1:10.

V. Discussion

Evolution of Jag UNG

Compared with the widely distributed family 1 UNG, the homologs of Jag UNG are only identified in *Oxalobacteraceae* family including the genus of *Janthinobacterium*, *Pseudoduganella, Massilia, Duganella, Noviherbaspirillum, Herbaspirillum and Collimonas.* Nsa UNG homologs are only found in the genus of *Nitratifractor* and *Sulfurovum,* both of which belong to the *Epsilonproteobacteria* family. Family 4 UDGa enzymes, existing in archaea, thermophilic bacteria and some bacteriophages, are wildly accepted as the common ancestor of other UDG families [18] . Interestingly, the organisms which contain the homologs of Jag UNG or Nsa UNG have a typical family 4 UDGa in their genome. Therefore, it is possible that gene duplication plays an important role in the evolution of the enzymes like Jag UNG and Nsa UNG. Following the gene duplication, the original family 4 UDGa still keeps the robust uracil cleavage activity and the duplicated one evolved to be family 1 UNG enzymes by point mutations, deletions and insertions.

Enzyme properties of Jag UNG

The exclusive but lower UDG activity of Jag UNG is closely related with its structure. It already gained the family 1-uniqued Tyr 79 that sterically blocks the C5 position of uracil that explained why Jag UNG only works on the uracil not hypoxanthine or xanthine. However, Jag UNG has not gain the catalytic residue Asp (D63 in *E.coli* UNG) in motif 1 of family 1, but has lost the catalytic residue Asn (N89 in TthUDGa) in

motif 3 of family 4 UDGa [25]. Therefore, its UDG activity is not as strong as family 1 UNG or family 4 UDGa.

EG to QD

The correlation of EG in the motif 1 of family 4 UDGa and the QD in the motif 1 of family 1 UNG has been confirmed previously [25]. In the present study, through the mutagenesis and kinetic study of Jag UNG, the potential evolutionary intermediate between family 4 UDGa and family 1 UNG, we further proved that E and G, Q and D are functionally correlated, respectively. Meanwhile, from the evolutionary view, we estimated that co-evolution from EG to QD is a key step in evolutionary path from family 4 UDGa to family 1 UNG. In the basis an evolutionary scenario of UDG superfamily, the highly active family 1 UNG was evolved from family 4 UDGa by acquiring the Asp in motif 1, which working as a general base to activate a water molecular [18]. According to our double mutagenesis data, family 1 UNG is evolved from the common ancestor by acquiring not only an Asp but a Gln. Naturally, Nsa UNG has already gained both Q and D and it has been closer to family 1 UNG and became a highly active UDG.

A82E

Structurally, A82 does not locate in the catalytic center, we proposed that replacement A82 with an acidic residue (E) might affect the overall structure of Jag UNG that could be responsible for the increase of enzyme efficiency of Jag UNG-A82E. MD simulation analysis demonstrated that A82E alters the root mean square fluctuation (RMSF) of the few regions in this protein. Among those regions, the region (79-86) is close to the catalytic center of Jag UNG (Fig. 4.7A). Notably, the salt bridge between

R81 and DNA backbone in this region plays an important role for the UDG activity of Jag UNG (Fig. 4.7B and 4.7C). Further analysis indicated that this salt bridge is much more stable in Jag UNG-A82E than that in Jag UNG-WT. As indicated in Fig. 4.7D, this salt bridge in Jag UNG-WT breaks after 25 ns while it still stable in Jag UNG-A82E. Meanwhile, the average hydrogen bond between A82E protein and DNA increased from 4.8 (WT) to 5.3 (Table 4.2). Both the stronger salt bridge and hydrogen bond may explain the lower k_m and higher enzyme efficiency of Jag UNG-A82E. From evolutionary view, losing the E and gain Y might be a necessary step from family 4 UDGa to family 1 UNG.

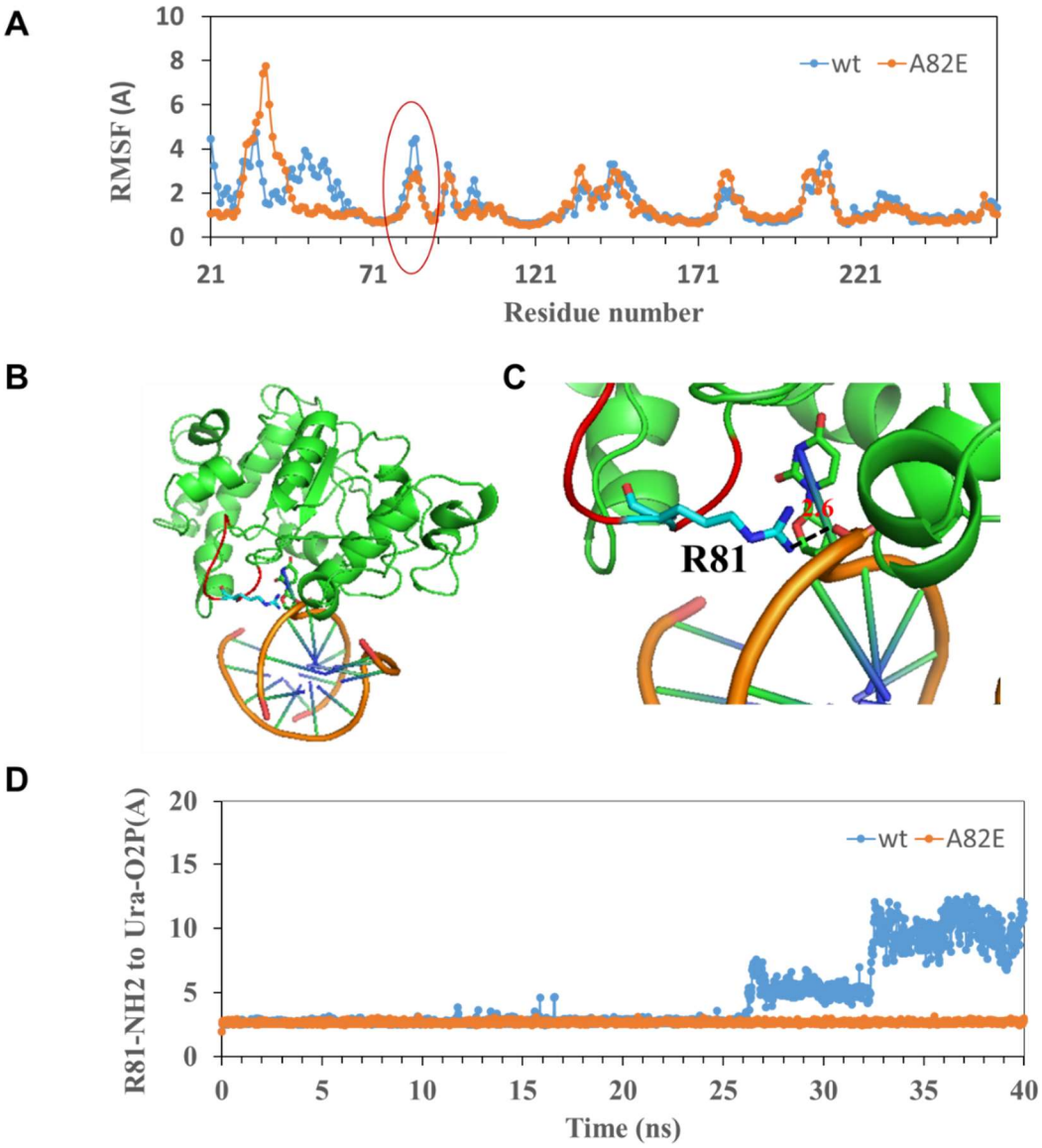


Figure 4. 7 A82E break the salt bridge between Jag UNG and DNA substrate. A. RMSF profiles for WT and A82E for Cα atoms of all residues in Jag UNG; B. The interactions between Jag UNG-R81 and modeled uracil-containing DNA (DNA model was taken from PDB 1EMH). Jag UNG structure is shown in cartoon mode and the residues R81 is shown in licorice; C. Close-up view of the interactions between Jag UNG-R81 and modeled uracil-containing DNA (DNA model was taken from PDB 1EMH). Jag UNG structure is shown in cartoon mode and the residue R81 is shown in licorice; D. The distance profile between R81-NH2 to Ura-O2P for WT and A82E from 0 to 40 ns in Jag UNG.

L245H

The increased enzyme efficiency of JagUNG-L245H could be explained by structural, electro potential and MD simulation analysis. Structurally, L245 locates near the interface between Jag UNG and uracil-containing DNA substrate. Replacement L245 with a positively charged H might increase the binding affinity of Jag UNG to the negatively charged DNA substrate by changing the electro potential. As showed in Fig. 4.8, the electrostatic potential of the mutation site (L245) change from negative to positive upon the mutation. Meanwhile, the MD simulation indicated that the average number of hydrogen bond between Jag UNG-L245H protein and uracil-containing DNA substrate increased from 4.8 (WT) to 5.2. Consistently, the Jag UNG-E76Q/G77D/L245H triple mutant averagely contains 6.2 hydrogen bonds between protein and DNA that is significantly higher than that of Jag UNG-E76Q/G77D double mutant (5.6) (Table 4.2). All those results are consistent with the higher binding affinity and enzyme efficiency of Jag UNG-L245H.

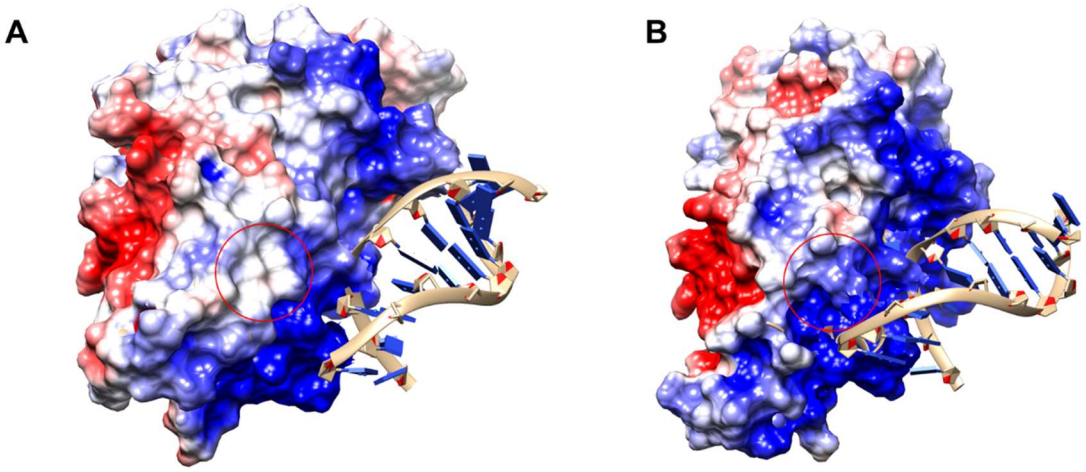


Figure 4. 8 Electrostatic potential of WT and L245H of Jag UNG. A. Electrostatic potential of the WT of Jag UNG; B. Electrostatic potential of L245H mutant of Jag UNG. The mutation site is marked with a red circle. The positive potential region is colored with bule and the negative potential region is colored with red.

Ugi effect

Ugi is a DNA mimic and small protein (80 aa), that was initially found in the PBS phase. It is in charge of protecting its uracil-containing genomic DNA from UDG in their host cell [27]. Ugi can form an extremely tight complex with the traditional family 1 UNG which is even stable in 8M urea [16]. By structural comparison, like family 4 UDGa and Nsa UNG, Jag UNG can't interact with Ugi as they have not involved the proper interface to bind Ugi. Notably, in the interface between family 1 UNG and Ugi, a longer loop is able to insert into the hydrophobic pocket of Ugi (Fig. 4.9D). However, both family 4 UDGa, Jag UNG and Nsa UNG only have a shorted loop that is not enough to interact with Ugi protein (Fig. 4.9A-9C). As a special interface is needed for the interaction between UDG and Ugi, it absolutely needs more complicated evolution to generate this elaborate interface.

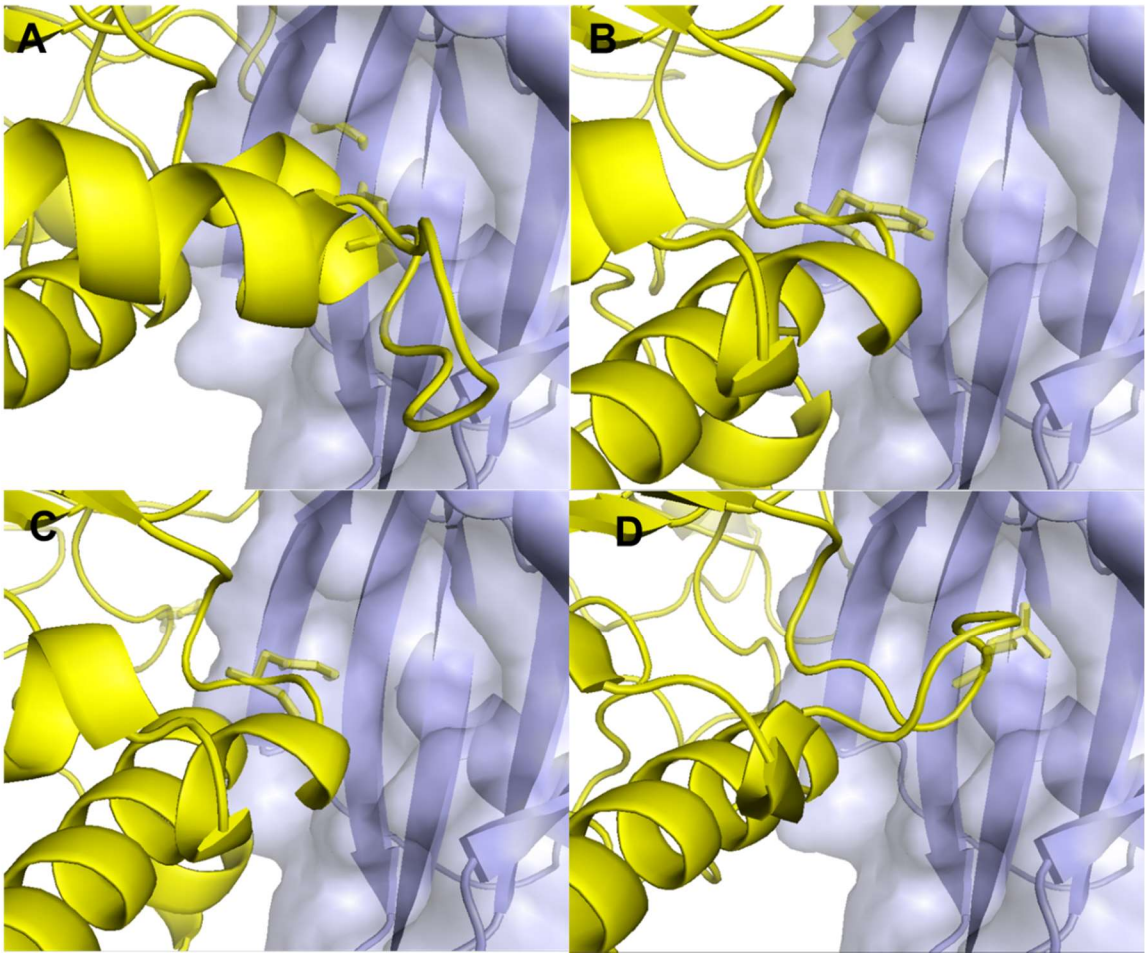


Figure 4. 9 Potential interactions between Jag UNG and Ugi and its comparison with Tth UDGa, Nsa UNG and human UNG. A. Modeled interactions between Tth UDGa and Ugi. (Ugi was taken from PDB 1UGH); B. Modeled interactions between Nsa UNG and Ugi. (Ugi was taken from PDB 1UGH); C. Modeled interactions between Jag UNG and Ugi. (Ugi was taken from PDB 1UGH); D. Interactions between human UNG and Ugi (PDB 1UGH).

In summary, we found a potential intermediate between family 1 UNG and family 4 UDGa, phylogenetic analysis attribute it to be a new subgroup of family 1 UNG. Through the enzymatic study and comparison with other family 1 UNG and family 4 UDGa enzymes, we partly illuminate the evolutionary mechanisms from family 4 UDGa to family 1 UNG.

VI. References

1. Parikh, S. S., Putnam, C. D. & Tainer, J. A. (2000) Lessons learned from structural results on uracil-DNA glycosylase, *Mutation research.* **460**, 183-99.

2. Krokan, H. E., Drablos, F. & Slupphaug, G. (2002) Uracil in DNA--occurrence, consequences and repair, *Oncogene.* **21**, 8935-48.

3. Cortazar, D., Kunz, C., Saito, Y., Steinacher, R. & Schar, P. (2007) The enigmatic thymine DNA glycosylase, *DNA repair.* **6**, 489-504.

4. Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J. & Schar, P. (2001) Thymine DNA glycosylase, *Prog Nucleic Acid Res Mol Biol.* **68**, 235-53.

5. Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily, *Mutation research.* **460**, 165-81.

6. Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W. & Verdine, G. L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors, *Current biology : CB.* **9**, 174-85.

7. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N. & Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG, *Journal of molecular biology.* **385**, 761-78.

8. Sandigursky, M. & Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes, *Current biology : CB.* **9**, 531-4.

9. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H. & Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site, *The EMBO journal.* **21**, 3182-91.

10. Lee, H. W., Dominy, B. N. & Cao, W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily, *The Journal of biological chemistry.* **286**, 31282-7.

11. Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S. & Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8, *Journal of molecular biology.* **333**, 515-26.

12. Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S. & Masui, R. (2007) Crystal structure of family 5 uracil-DNA glycosylase bound to DNA, *Journal of molecular biology.* **373**, 839-50.

13. Wang, H. C., Ho, C. H., Chou, C. C., Ko, T. P., Huang, M. F., Hsu, K. C. & Wang, A. H. (2016) Using structural-based protein engineering to modulate the differential inhibition effects of SAUGI on human and HSV uracil DNA glycosylase, *Nucleic acids research.* **44**, 4440-9.

14. Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X. & Cheng, X. (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation, *Nucleic acids research.* **40**, 10203-14.

15. Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L. & Pearl, L. H. (2003) Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1, *Mol Cell.* **11**, 1647-59.

16. Acharya, N., Kumar, P. & Varshney, U. (2003) Complexes of the uracil-DNA glycosylase inhibitor protein, Ugi, with Mycobacterium smegmatis and Mycobacterium tuberculosis uracil-DNA glycosylases, *Microbiology.* **149**, 1647-58.

17. Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA, *Cell.* **82**, 701-8.

18. Aravind, L. & Koonin, E. V. (2000) The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates, *Genome Biol.* **1**, RESEARCH0007.

19. Fisher, C. L. & Pei, G. K. (1997) Modification of a PCR-based site-directed mutagenesis method, *BioTechniques.* **23**, 570-1, 574.

20. King, K., Benkovic, S. J. & Modrich, P. (1989) Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease, *The Journal of biological chemistry.* **264**, 11807-15.

21. Vermote, C. L. & Halford, S. E. (1992) EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition, *Biochemistry.* **31**, 6082-9.

22. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics.* **22**, 195-201.

23. Khoury, G. A., Tamamis, P., Pinnaduwage, N., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. (2014) Princeton_TIGRESS: protein geometry refinement using simulations and support vector machines, *Proteins.* **82**, 794-814.

24. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem.* **25**, 1605-12.

25. Xia, B., Liu, Y., Guevara, J., Li, J., Jilich, C., Yang, Y., Wang, L., Dominy, B. N. & Cao, W. (2017) Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily, *Sci Rep.* **7**, 45978.

26. Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A. & Krokan, H. E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase, *EMBO J.* **15**, 3442-7.

27. Cone, R., Bonura, T. & Friedberg, E. C. (1980) Inhibitor of uracil-DNA glycosylase induced by bacteriophage PBS2. Purification and preliminary characterization, *The Journal of biological chemistry.* **255**, 10354-8.

CHAPTER FIVE

RESEARCH SIGNIFICANCE AND CONCLUSION REMARKS

UDG enzymes are essential factor in the BER pathway and play crucial role in removing deaminated DNA bases. In UDG superfamily, each family exhibit distinct structural and catalytic features. In this thesis, I identified three different UDG enzymes which are classified as new groups of existing family by sequence alignment and phylogenetic analysis. Meanwhile, biochemical, structural, bioinformatics and biophysical studies were performed to identify their special catalytic mechanism.

Even though the traditional Family 3 UDG is termed as single-strand-selective monofunctional uracil DNA glycosylase (SMUG1), it works on both single-stranded and double-stranded uracil-containing DNA substrate [1-5]. In Chapter 2, I characterized a SMUG1-like enzyme from *Listeria Innocua* which only works on single-stranded DNA substrate at various conditions. This single-stranded UDG activity of Lin SMUG1-like is consistent with the lower G/C content in the host genome which tend to form single-stranded DNA. Comparison of with the traditional SMUG1 enzymes, SMUG1-like has a similar catalytic center, except SMUG1-like contains double Ser while SMUG1 adapt Met and Asn in their catalytic center. Mutagenesis and mutual information analysis demonstrated that double Ser in SMUG1-like and Met and Asn in SMUG1 are functional correlated. This is the first time to report the correlated residues in family 3 SMUG1 enzymes.

In chapter 3, I characterized another novel UDG enzyme from *Nitratifractor salsuginis* (Nsa UNG) which belongs to a new subfamily of family 1 UNG. Similar with

classic family 1 UNG, Nsa UNG exhibit extremely robust and exclusive UDG activity [6]. However, Nsa UNG harbors two distinct features. In one way, Nsa UNG form a novel salt bridge network with uracil-containing DNA backbone. Any disruption of this salt bridge network inactivate this enzyme. This newly discovered salt bridge network may contributes the lower K_m of Nsa UNG to uracil-containing DNA. In another way, unlike traditional family 1 UNG, Nsa UNG can't be inhibited by Ugi. Structural analysis showed that Nsa UNG doesn't evolved a proper interface for Ugi binding. Also, the binding energy between Nsa UNG and Ugi is larger than that of traditional UNG and Ugi. The discovery of this enzyme broaden family 1 UNG and further verified the diverse substrate recognition and catalytic mechanism of UDG superfamily.

In chapter 4, another UDG from *Janthinobacterium agaricidamnosum* (Jag UNG) was studied. In the functional motifs, Jag UNG contains both features of family 1 UNG (including Nsa UNG) and family 4 UDGa. Even though the sequence alignment and phylogenetic analysis classified Jag UNG as another new group of family 1 UNG, Jag UNG only exhibit very limit UDG activity as lacking some catalytic residues. Gaining the catalytic residues Q76 and D77 make this enzyme much stronger, which is consistent the correlation study of Q and D in motif 1 of family 1 UNG previously [7]. In addition, we found two mutations make this enzyme stronger. In motif 1, introducing an A82E in motif 1 stabilizes the salt bridge between Jag UNG and DNA, thus increase the binding affinity and enzyme efficiency. In motif 2, L245H also elevate the binding affinity and enzyme efficiency by changing the electro potential of the interface between enzyme and

DNA. The discovery of this enzyme make us deeper understand the evolutionary and catalytic mechanism of UDG superfamily.

In summary, this thesis described three different UDG enzymes and contribute to the pedigree of UDG superfamily. Identification the novel catalytic residues, salt bridge network, and elevate mutant proved that the catalytic mechanism of UDG enzyme is much more diverse and complicate. In addition, the discovery of correlated residues make us have a better understanding of the evolutionary mechanism of UDG superfamily. In the future, with the discovery of more and more members of UDG enzymes, the evolutionary processing and the diverse catalytic mechanism of UDG superfamily will be better illuminated.

Besides this the projects included in this thesis, I participated the study of other UDGs and MeCP2 which resulted in four peer-reviewed publications:

1. Bo Xia, Yinling Liu, Jose Guevara, **Jing Li**, Celeste Jilich, Liangjiang Wang, Brian N. Dominy, Weiguo Cao. Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily. *Scientific reports*. 2017, 7:45978

2. Panjiao Pang, Ye Yang, **Jing Li**, Wang Z, Weiguo Cao, Wei Xie. SMUG2 DNA Glycosylase from *Pedobacter heparinus* as a New Subfamily in UDG Superfamily. *Biochem J.* 2017, 474(6):923-938

3. Zhemin Zhang, Jiemin Shen, Ye Yang, **Jing Li**, Weiguo Cao, Wei Xie. Structural basis of substrate specificity in *Geobacter metallireducens* SMUG1. *ACS Chem Biol.* 2016, 11(6):1729-36.

4. Ye Yang*, Kucukkal TG*, **Jing Li**, Emil Alexov, Weiguo Cao. Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. *ACS Chem Biol.* 2016, 11(10):2706-2715. *, These authors contribute equally

Reference

1. Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L. & Lindahl, T. (2001) Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase, *EMBO J.* **20**, 4278-86.

2. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E. & Slupphaug, G. (2002) hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup, *J Biol Chem.* **277**, 39926-36.

3. Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E. & Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms, *Nucleic acids research.* **35**, 3879-92.

4. Matsubara, M., Tanaka, T., Terato, H., Ohmae, E., Izumi, S., Katayanagi, K. & Ide, H. (2004) Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase, *Nucleic acids research.* **32**, 5291-302.

5. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N. & Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG, *Journal of molecular biology.* **385**, 761-78.

6. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues, *Proceedings of the National Academy of Sciences of the United States of America.* **71**, 3649-53.

7. Xia, B., Liu, Y., Guevara, J., Li, J., Jilich, C., Yang, Y., Wang, L., Dominy, B. N. & Cao, W. (2017) Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily, *Sci Rep.* **7**, 45978.